TIGHT BINDING BOOK

OU_158310

THEORY AND APPLICATION
OF MICROWAVES

Theory and Application

of Microwaves

BY

ARTHUR B. BRONWELL, M.S., M.B.A.

President, Worcester Polytechnic Institute
Formerly Professor of Electrical Engineering
Northwestern University
Secretary, American Society for Engineering Education

ROBERT E. BEAM, Ph.D.

Professor of Electrical Engineering
Northwestern University

THEORY AND APPLICATION OF MICROWAVES

Preface

In this book, the authors have endeavored to present the underlying theory of microwave systems, starting with concepts which are basic to all electromagnetic phenomena. The material included in the text logically falls into three general categories: (1) fundamental electronic concepts and their application in the analysis of microwave tubes, (2) transmission lines and transmission-line networks, and (3) electromagnetic field equations and their use in the analysis of wave propagation, reflection phenomena, wave guides, and radiating systems. Throughout the book the engineering point of view has been stressed and, wherever possible, the analytical results have been expressed in a form convenient for engineering use.

In recent years, there has been a growing trend toward the development of new and unorthodox types of vacuum tubes, many of which have been designed for operation at microwave frequencies. This trend has emphasized the need for enlarging upon and clarifying our fundamental electronic concepts, particularly those dealing with space-charge behavior in time-varying fields. The groundwork in this subject has been admirably treated by Benham, Llewellyn, North, Jen, Gabor, and others. However, the mathematical complexities involved in a rigorous treatment of the subject are such as to discourage all those who are not endowed with an abundance of mathematical fortitude. In this text, the authors have attempted to present some of the fundamental electronic concepts in simplified form. Wherever necessary, rigor has been sacrificed in the interests of clarity and conciseness. There follow several chapters dealing with the theory and description of microwave tubes, including the triode, klystron, reflex klystron, resnatron, and magnetron.

The theory of transmission lines and the use of impedance diagrams in the solution of transmission-line problems are given in Chaps. 8 and 9. Emphasis has been placed upon the use of P. H. Smith's polar impedance diagram, since this has proved most useful in engineering practice. A slight modification of this impedance diagram has been made in order to facilitate the solution of transmission-line problems in which the lines have losses. The following chapter on transmission-line networks is reasonably com-

plete and includes certain aspects of microwave networks which heretofore have not been published in textbook form.

A broad survey of typical microwave systems is given in Chaps. 11 and 12. No attempt has been made to treat these systems exhaustively. The greatest emphasis has been placed upon the microwave portions of the systems, since there are a number of excellent radio textbooks which adequately cover the aspects dealing with radio-circuit theory.

Maxwell's equations are introduced in Chap. 13 and are followed by a chapter on wave propagation and reflection. The solution of the wave equation in various coordinate systems is treated in Chap. 15. Subsequent chapters deal with the analysis of wave guides, resonators, horns, and antenna systems.

Rationalized mks units have been used throughout the text.

This book is an outgrowth of courses taught to senior and first-year graduate students in electrical engineering at Northwestern University for the past six years. Several factors influenced the arrangement of the material. The chapters dealing with fundamental electronic concepts and microwave tubes have been placed at the beginning of the book in order to facilitate an early start in the laboratory work. It has also been found that this arrangement provides a convenient review of field concepts before taking up Maxwell's equations.

In teaching senior courses, selected portions of the text may be used. One arrangement might include Chaps. 1 to 3, 6 to 10, 13, 15, 16, 18, and 19. These chapters have been written in somewhat simplified form for this purpose. For those who prefer to concentrate their efforts on field theory, Chaps. 1 and 2, together with the last half of the book starting with Chap. 13, are recommended. In general, a knowledge of radio-circuit theory and mathematics through calculus will be required for an understanding of the analytical portions of the text.

The authors have drawn freely from the published works of many scientists and engineers who have contributed to the present-day knowledge of this subject. They wish to express their grateful appreciation for the assistance received from these sources. Special appreciation is extended to Dean O. W. Eshbach and to Dr. J. F. Calvert for their helpful encouragement and to Northwestern University for its liberal policy which made this work possible. Generous assistance was also received from the Radiation Laboratory of the Massachusetts Institute of Technology, Sperry Gyroscope Co., Bell Telephone Laboratories, General Electric Co., and others. Mr. P. H. Smith kindly permitted the use of his polar impedance diagram. The authors are indebted to Nirmal Mondol for his assistance in preparing the illustrations. Finally, but by no means least, they wish to express their grateful appreciation to Hope Beam and Virginia Bronwell

for their constant encouragement in the preparation of the manuscript and for their generous assistance in reading the proof.

<div align="right">

ARTHUR B. BRONWELL
ROBERT E. BEAM

</div>

Evanston, Ill.,
   August, 1947.

Contents

CHAPTER 1

INTRODUCTION

CHAPTER 2

CHARGES IN ELECTRIC FIELDS

CHAPTER 3

CURRENT, POWER, AND ENERGY RELATIONSHIPS

CHAPTER 4

THE PHYSICAL BASIS OF EQUIVALENT CIRCUITS

CHAPTER 5

NEGATIVE-GRID TRIODE OSCILLATORS AND AMPLIFIERS

CHAPTER 6

TRANSIT-TIME OSCILLATORS

CHAPTER 7

MAGNETRON OSCILLATORS

<div align="center">CHAPTER 8</div>

<div align="center">*TRANSMISSION-LINE EQUATIONS*</div>

<div align="center">CHAPTER 9</div>

<div align="center">*GRAPHICAL SOLUTION OF TRANSMISSION-LINE
PROBLEMS*</div>

<div align="center">CHAPTER 10</div>

<div align="center">*TRANSMISSION-LINE NETWORKS*</div>

CHAPTER 11

TRANSMITTING AND RECEIVING SYSTEMS

CHAPTER 12

PULSED SYSTEMS—RADAR

CHAPTER 13

MAXWELL'S EQUATIONS

CHAPTER 14

PROPAGATION AND REFLECTION OF PLANE WAVES

CHAPTER 15

SOLUTION OF ELECTROMAGNETIC-FIELD PROBLEMS

CHAPTER 16

WAVE GUIDES

CHAPTER 17

*IMPEDANCE DISCONTINUITIES IN GUIDES—
RESONATORS*

CHAPTER 18

*APPLICATIONS OF WAVE GUIDES AND
RESONATORS*

CHAPTER 19

LINEAR ANTENNAS AND ARRAYS

CHAPTER 20

IMPEDANCE OF ANTENNAS

CHAPTER 21

OTHER RADIATING SYSTEMS

CHAPTER 1

INTRODUCTION

The history of the exploration and utilization of new and untried areas of science follows a strangely uniform evolutionary pattern. The fundamental physical laws are first discovered by theoretical and experimental scientists who piece together the fragmentary evidence into a coherent and integrated theory which opens the door to further development. This is often followed by years of slow and painstaking progress devoted to the discovery and refinement of new experimental techniques and analytical methods. Once the commercial possibilities of the new science become generally recognized, there follows a period of intensive engineering research and development, during which the mysteries of the science are transformed into everyday engineering principles. Such has been the pattern of research and development of microwaves.

1.01. Historical Developments.—In 1864, Maxwell laid the foundation of our modern concepts of electromagnetic theory, a theory which he used to explain the phenomenon of light-wave propagation. Scarcely 25 years later, Hertz, experimenting with a spark transmitter, generated and received electromagnetic waves having wavelengths of the order of 60 centimeters. These experiments were performed in order to confirm the existence of the electromagnetic waves predicted by Maxwell. Within a decade, wavelengths as short as 0.4 centimeter had been realized by other experimenters using similar methods.

The vastly superior characteristics of the conventional vacuum tube, at longer wavelengths, ushered in an era of rapid scientific, engineering, and commercial development of radio. However, serious difficulties were soon encountered in attempting to extend the range of operation of the vacuum tube to the shorter wavelengths. As the wavelength approached the order of magnitude of the physical dimensions of the tube, such mystifying properties as electron transit time, interelectrode capacitance and conductance, and lead inductance appeared to impose definite lower limits to the wavelengths which could be generated in the conventional types of tubes. It became apparent that a new approach was necessary. The pioneering work of Barkhausen on the positive-grid oscillator, Hull and others on the magnetron, and later the Varian brothers on the klystron, broke away from traditional ideas to develop fundamentally new types of vacuum tubes for the generation of microwaves.

1

While some experimenters were moving in the direction of ever shorter wavelengths in the microwave spectrum, others were finding methods of generating longer infrared wavelengths. Eventually the union occurred and, at present, wavelengths in the infrared spectrum have been generated in electron tubes.

In the development of microwave networks, it became necessary to abandon the conventional lumped inductances and capacitances and turn to distributed parameter systems such as transmission lines, wave guides, and hollow metallic resonators. The theory of electromagnetic wave propagation through hollow conducting tubes (wave guides) was first presented by Rayleigh in 1897 but lay relatively dormant for over 30 years, pending the development of microwave generators. The relatively recent theoretical and experimental work of Southworth, Barrow, and others demonstrated not only the physical realizability of wave propagation through wave guides, but also that, under certain conditions, such systems may be actually superior to the two-conductor transmission line for the transmission of microwave energy.

1.02. Generation of Light Waves and Radio Waves.—Electromagnetic waves are generated by electric charges moving through retarding electromagnetic fields. The charge is decelerated by the field, thereby giving up part of its energy to the field. Under the proper circumstances, the energy released by the charge can appear as an electromagnetic wave in space.

Let us briefly compare the generation of light waves and radio waves. According to the Bohr theory, the atom consists of a positively charged nucleus and negatively charged electrons which rotate about the nucleus in elliptical orbits, each orbit corresponding to a definite energy level. Owing to atomic collisions or other causes, an electron may momentarily be displaced to an unstable higher energy level. In returning to a lower energy level, the electron is retarded by the electric field of the atom, thereby releasing a discrete amount of energy which appears in the form of electromagnetic radiation. The amount of energy released determines the wavelength of the radiation. Each electron transition is the source of an electromagnetic wavelet; hence the generation of light waves is a random phenomenon.

In contrast with the random oscillations of light waves, radio-frequency oscillations can be generated by an orderly, controlled stream of electrons moving in the electric field of a vacuum tube. The electrons are accelerated by a d-c electric field and take energy from the potential source which produces this field. They pass through an alternating field in such a phase as to be retarded by this field and, hence, give up a portion of their energy to the alternating field. The source of the alternating field usually consists of some form of oscillatory circuit which receives the cumulative

energy given up by a large number of electrons and sets up the necessary alternating field in the vacuum tube.

The electromagnetic energy may then pass through an elaborate system of electric circuits before it is ultimately radiated into space or dissipated. Many of these circuits do not have their counterpart in the light spectrum. For example, in the generation of light waves, there is no known means of causing a large number of electrons in different atoms to oscillate in exact synchronism in a manner similar to that obtained in certain types of vacuum tubes or in oscillating electric circuits. On the other hand, both radio waves and light waves possess similar properties of reflection, refraction, diffraction, polarization, and interference. Thus it is possible to construct the microwave counterparts of optical filters, lenses, reflectors, diffraction gratings, spectrographs, and interferometers.

As the microwave developments enter the region bordering on the infrared spectrum, we should expect the techniques of generation, control, and utilization of microwaves to take on the common aspects of both fields. Experience gained in either field will inevitably aid in a better understanding of the other.

1.03. Engineering Considerations.—In this text we are primarily concerned with the theoretical and practical aspects of vacuum tubes, systems, and radiation phenomena in the frequency range from 300 to 300,000 megacycles, corresponding to a wavelength range of 1 meter to 1 millimeter. Since it is helpful to have an over-all term to describe this portion of the radiation spectrum, these frequencies will arbitrarily be referred to as *microwave frequencies.*

From an engineering point of view, the commercial development of microwave systems opens up enormous new areas for radio broadcasting, television, radar, radio relaying, navigational systems, and special services. A unique feature of microwave systems is the ease of obtaining highly directional radiation. Because of the short wavelengths, antennas are physically small so that parabolic reflectors or multielement directional arrays can be used conveniently. For this reason, the microwave portion of the frequency spectrum is especially well suited to radar, navigational systems, point-to-point communication, and radio relaying systems.

Microwaves are seldom, if ever, reflected from the ionosphere layers. Consequently the maximum useful range of transmission is limited to horizon distances. For land-based systems, this distance is of the order of 25 to 200 miles, depending upon the height of the transmitting and receiving antennas. Longer ranges can be obtained with air-borne systems. Experiments have shown that, when certain exceptional atmospheric conditions prevail, microwave reflections can occur from the troposphere, which is the portion of the earth's atmosphere below the ionosphere.

These reflections make it possible, occasionally, to transmit and receive signals beyond horizon distances.

Basically, the communication systems for microwave frequencies contain the same functional components as those operating at ordinary radio frequencies. Thus, the transmitting system may consist of a modulated oscillator and some form of radiating system. Superheterodyne receivers are often used in microwave communication systems. A typical receiving system would contain the receiving antenna, a local oscillator and converter, several stages of intermediate-frequency amplification, a detector, and one or more stages of audio or video amplification.

Although the basic systems are the same as those used at radio frequencies, the vacuum tubes and circuit elements of microwave systems are quite different. At microwave frequencies, we are likely to find klystrons, magnetrons, resnatrons, or other types of tubes designed specifically for operation at these frequencies. The networks are comprised of transmission-line elements, wave guides, cavity resonators, and other components which are quite foreign to conventional radio-frequency systems.

1.04. Theoretical Aspects.—In our analysis of microwave tubes and circuits, we shall find it necessary to abandon some of the physical and analytical concepts which have served quite satisfactorily as approximations at lower frequencies but which in some cases are not sufficiently fundamental to account correctly for the microwave phenomena. For example, in the analysis of vacuum-tube circuits it has become customary to represent the tube by an equivalent circuit, thereby reducing an electronic problem to the status of an electric-circuit problem. While this expedient greatly simplifies the analysis, it also serves to conceal the fundamental physical phenomena taking place inside the tube.

A more fundamental approach is to start with the equations of motion of electrons in electric and magnetic fields and use them to determine the behavior of the electrons in the tube as well as their external electrical effects. Although this is a fundamental method of analysis which is valid for all types of tubes, regardless of the operating frequency, it often involves serious mathematical complications which restrict its use. In this text, the fundamental method of analysis is presented as a means of obtaining a better understanding of the true physical processes at work. The analysis is applied to systems of relatively simple geometry, although the physical concepts are valid for all possible systems.

We shall find a close resemblance between electromagnetic-wave propagation along transmission lines, in wave guides, and in free space. In studying wave propagation along transmission lines and in wave guides, it becomes increasingly important to adopt a field viewpoint based upon Maxwell's equations rather than the conventional circuit viewpoint. The dielectric surrounding the transmission line or inside the wave guide is

considered to be the locale of energy storage and energy flow. The conductors merely serve to guide the energy flow in the dielectric while, at the same time, exacting their toll in the form of I^2R loss due to currents in the conductors.

Transmission lines may be analyzed either by using the circuit method based upon Kirchhoff's laws for voltages and currents or by solving Maxwell's field equations subject to the given boundary conditions. The former method of analysis involves a one-dimensional problem and hence is considerably easier than the three-dimensional solution of the field equations. However, the Kirchhoff-law solution fails to show the existence of important "higher modes" which are likely to occur at microwave frequencies. These higher modes have an entirely different field distribution in space from that existing at lower frequencies. The circuit method of analysis will be considered first in this text, since it offers the simplest solution for most engineering problems, including those at microwave frequencies. The higher order modes will be discussed in the chapter on wave guides.

CHAPTER 2

CHARGES IN ELECTRIC FIELDS

Basically, all vacuum tubes employ the effects of electrons moving under the influence of electric and magnetic fields. The principles of electron dynamics therefore represent the foundation upon which we can safely build our physical and analytical concepts. We shall find that there is an intimate relationship between the dynamical behavior of the electrons and the electrical quantities.

FIG. 1.—Representation of vector quantities in rectangular coordinates.

In this chapter we shall consider the laws governing the electric field distribution in space as well as the equations of motion of charged particles in static and time-varying electric fields. Later chapters deal with the relationships between the dynamical behavior of the electrons and the electrical quantities.

2.01. Vector Manipulation.[1,2]— Since electromagnetic theory deals with fields in three-dimensional space, vector analysis is a logical system of mathematical expression. Let us, therefore, briefly consider some properties of vector and scalar quantities.

A vector is a quantity which has both direction and magnitude. A scalar quantity has only magnitude. Thus force, distance, electric intensity, and flux density are vector quantities. On the other hand, power, energy, potential, and total flux are scalar quantities.

Any vector may be represented as the sum of three mutually orthogonal component vectors. In rectangular coordinates, a vector $\bar{A}$, shown in Fig. 1, is represented as

$$\bar{A} = A_x \bar{\imath} + A_y \bar{\jmath} + A_z \bar{k} \tag{1}$$

where A_x, A_y, and A_z are the scalar magnitudes of the components of $\bar{A}$ along the x, y, and z axes, respectively, and $\bar{\imath}$, $\bar{\jmath}$, and $\bar{k}$ are unit vectors in these respective directions. A unit vector serves to assign a direction to

[1] PHILLIPS, H. B., "Vector Analysis," John Wiley & Sons, Inc., New York, 1933.

[2] COFFIN, J. G., "Vector Analysis," John Wiley & Sons, Inc., New York, 1938.

the quantity by which it is multiplied. The scalar magnitude of vector $\bar{A}$ is

$$A = \sqrt{A_x^2 + A_y^2 + A_z^2} \tag{2}$$

Consider the two vectors $\bar{A}$ and $\bar{B}$, where

$$\bar{A} = A_x \bar{\imath} + A_y \bar{\jmath} + A_z \bar{k}$$

$$\bar{B} = B_x \bar{\imath} + B_y \bar{\jmath} + B_z \bar{k} \tag{3}$$

These vectors may be added or subtracted by adding or subtracting their components as follows:

$$\bar{A} + \bar{B} = (A_x + B_x)\bar{\imath} + (A_y + B_y)\bar{\jmath} + (A_z + B_z)\bar{k}$$

$$\bar{A} - \bar{B} = (A_x - B_x)\bar{\imath} + (A_y - B_y)\bar{\jmath} + (A_z - B_z)\bar{k} \tag{4}$$

There are two types of vector multiplication. These are known as the *dot product* (or scalar product) and the *cross product* (or vector product). The dot product of two vectors $\bar{A} \cdot \bar{B}$ is defined by

$$\bar{A} \cdot \bar{B} = AB \cos \theta_{AB} \tag{5}$$

where A and B are the scalar magnitudes of vectors $\bar{A}$ and $\bar{B}$, and θ_{AB} is the angle between them. The dot product of two vectors is a scalar quantity.

Since the unit vectors $\bar{\imath}$, $\bar{\jmath}$, and $\bar{k}$ all have unit length, Eq. (5) shows that the dot product of two unit vectors in the same direction is unity, while the dot product of two unit vectors which are perpendicular is zero.

The dot product of vectors $\bar{A}$ and $\bar{B}$ in Eq. (3) may be obtained by taking the dot product of each component of $\bar{A}$ with every component of $\bar{B}$. This yields the scalar quantity

$$\bar{A} \cdot \bar{B} = A_x B_x + A_y B_y + A_z B_z \tag{6}$$

The cross product of two vectors designated $\bar{A} \times \bar{B}$ is defined by the relation

$$\bar{A} \times \bar{B} = AB \sin \theta_{AB} \bar{n} \tag{7}$$

where A and B are again the scalar magnitudes and θ_{AB} is the smaller of the two angles between the vectors. The cross product yields a vector which is perpendicular to both $\bar{A}$ and $\bar{B}$, this being the direction of the unit vector $\bar{n}$ in Eq. (7).

The direction of the vector representing the cross product may be determined by the right-hand rule. If we point the fingers of the right hand in the direction of vector $\bar{A}$, and close the fingers in the direction from vector $\bar{A}$ to vector $\bar{B}$ through the angle θ_{AB}, the extended thumb points in the direction of the vector representing $\bar{A} \times \bar{B}$. This direction is the same

as the direction of advance of a right-handed screw when turned in the direction from $\bar{A}$ to $\bar{B}$. It is apparent from this rule that $\bar{A} \times \bar{B} = -\bar{B} \times \bar{A}$, *i.e.*, the commutative law of multiplication does not apply to the cross product.

If Eq. (7) is applied to the unit vectors, we find that the cross product of two unit vectors pointed in the same direction is zero. The cross product of two unit vectors which are mutually perpendicular gives the third unit vector, its sign being determined by the right-hand rule. The sign of the cross product of two unit vectors can be obtained by writing the unit vectors in the triangular form

$$\underset{\bar{\underline{k}} \quad \bar{\jmath}}{\overset{\bar{\imath}}{\Big)}}$$
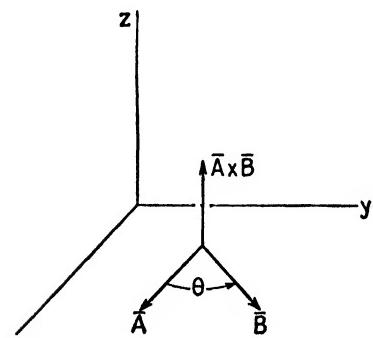
FIG. 2.—Cross-product multiplication.

Taking the vectors in the order in which they appear in the cross product, a positive sign is used for clockwise rotation in the above diagram and a negative sign for counterclockwise rotation. Thus, $\bar{\imath} \times \bar{\jmath} = \bar{\underline{k}}$, whereas $\bar{\jmath} \times \bar{\imath} = -\bar{\underline{k}}$.

The term-by-term cross product of the two vectors $\bar{A}$ and $\bar{B}$ yields

$$\bar{A} \times \bar{B} = (A_y B_z - A_z B_y)\bar{\underline{\imath}} + (A_z B_x - A_x B_z)\bar{\jmath} + (A_x B_y - A_y B_x)\bar{\underline{k}} \quad (8)$$

This result is also given by the expansion of the determinant

$$\bar{A} \times \bar{B} = \begin{vmatrix} \bar{\imath} & \bar{\jmath} & \bar{\underline{k}} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix} \quad (9)$$

Vector quantities may be expressed in cylindrical and spherical coordinates as well as in rectangular coordinates. In cylindrical coordinates, the coordinates of a point are ρ, ϕ, and z, as shown in Fig. 3, and the vector $\bar{A}$ is represented by

$$\bar{A} = A_\rho \bar{\rho} + A_\phi \bar{\phi} + A_z \bar{\underline{k}} \quad (10)$$

where $\bar{\rho}$, $\bar{\phi}$, and $\bar{\underline{k}}$ are unit vectors in the directions of increase of ρ, ϕ, and z, respectively. The differential volume $d\tau$ in cylindrical coordinates is $d\tau = \rho \, d\rho \, d\phi \, dz$.

In spherical coordinates a vector is represented by r, θ, and ϕ components as shown in Fig. 4, thus

$$\bar{A} = A_r \bar{\underline{r}} + A_\theta \bar{\theta} + A_\phi \bar{\phi} \quad (11)$$

where $\bar{r}$, $\bar{\theta}$, and $\bar{\phi}$ are unit vectors. The differential volume is $d\tau = r^2 \sin \theta \, dr \, d\theta \, d\phi$. Equations for the conversion from cylindrical and spherical coordinates to rectangular coordinates are given in Appendix III.

2.02. Coulomb's Law, Electric Intensity, and Potential in Electrostatic Fields.—Coulomb's law states that the force of attraction between two point charges is directly proportional to the product of the charges and
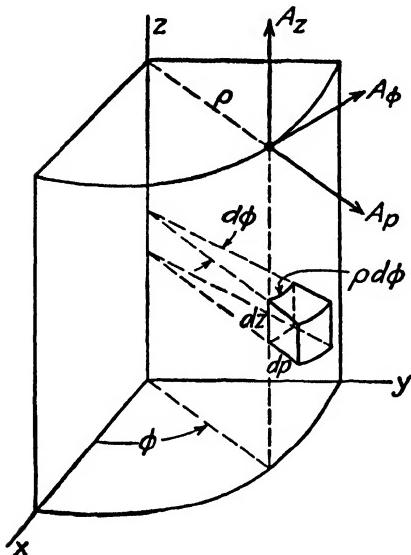
FIG. 3.—Cylindrical coordinates.   FIG. 4.—Spherical coordinates.
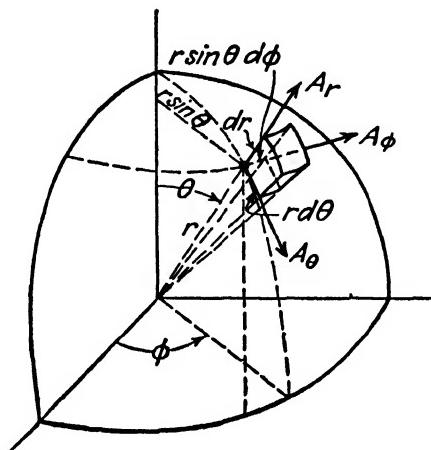
inversely proportional to the square of the distance between them. Thus, two charges q_1 and q_2 which are separated by a distance r experience a force

$$\bar{f} = \frac{kq_1q_2}{\epsilon r^2} \bar{r} \tag{1}$$

where ϵ is the permittivity of the medium, k is a proportionality constant, and $\bar{r}$ is a unit vector in the direction of the force.

In this text we shall use a rationalized mks (meter-kilogram-second) system of units.[1] In rationalized units the term 4π is included in the permittivity ϵ and permeability μ so that it does not appear in the principal electromagnetic field equations with which we will be dealing. However, in this system of units the 4π factor reappears in Coulomb's law and in equations in which electric intensity and potential are expressed in terms of the charges producing the field. Thus, in rationalized units Coulomb's

---

[1] A discussion of systems of units is given in Appendix I.

law becomes

$$\bar{f} = \frac{q_1 q_2}{4\pi\epsilon r^2}\,\bar{r} \tag{2}$$

In mks units, force is in newtons, charge in coulombs, and distance in meters.

The permittivity ϵ may be expressed as the product of the permittivity of free space ϵ_0 and a relative permittivity ϵ_r, thus

$$\epsilon = \epsilon_r \epsilon_0 \tag{3}$$

In rationalized mks units we have $\epsilon_0 = 1/(4\pi \times 9 \times 10^9) = 8.854 \times 10^{-12}$ farads per meter. The relative permittivity is the familiar dielectric constant, values of which are found in handbook tables.

Electric intensity is defined as the force on a unit positive charge placed in the field (assuming that the unit charge does not disturb the underlying field). Therefore the electric intensity at a point distant r from a point charge q may be found by setting q_2 equal to unity in Eq. (2), yielding

$$\bar{E} = \frac{q}{4\pi\epsilon r^2}\,\bar{r} \tag{4}$$
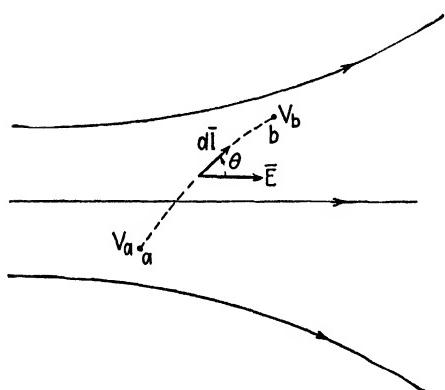
Fig. 5.—Potential difference is the line integral of electric intensity.

The mks unit of electric intensity is the volt per meter. In general, the field at a particular point may be due to a large number of charges distributed throughout space.

In an electrostatic field, the potential at a point in space is defined as the work done in moving unit charge against the forces of the field from a point of zero potential (sometimes assumed to be at infinity) to the point in question. In practical problems, we are usually more concerned with potential differences. Thus, in Fig. 5 the potential difference $V_b - V_a$ between two points a and b is the work done in carrying unit charge from point a to point b. Work may be expressed as the integral of the component of force in the direction of travel times distance of travel, or $w = \int \bar{f} \cdot d\bar{l}$. The force required to overcome the field is $\bar{f} = -\bar{E}$, hence the potential difference becomes

$$V_b - V_a = -\int_a^b \bar{E} \cdot d\bar{l} \tag{5}$$

An integral such as that of Eq. (5) is known as a *line integral*. The potential difference between two points, therefore, is the negative line integral of electric intensity along a path connecting the two points. Potential is sometimes written as an indefinite integral, thus,

$$V = -\int \bar{E} \cdot d\bar{l} \tag{6}$$

The mks unit of electric potential is the volt.

The potential V at a point distant r from a point charge q may be obtained by inserting Eq. (4) into (6). The integration constant is evaluated by assuming that the potential is zero at $r = \infty$, giving

$$V = \frac{q}{4\pi\epsilon r} \tag{7}$$

The potential at a point due to a number of discrete charges is the algebraic sum of the potentials resulting from the individual charges, or for n charges, we have

$$V = \sum_{i=1}^{i=n} \frac{q_i}{4\pi\epsilon r_i} \tag{8}$$

2.03. Potential Gradient.—Equation (2.02-5) expresses a relationship between potential and electric intensity in integral form. We shall now obtain a differential equation relating these quantities. The potential difference dV between two points an infinitesimal distance dl apart may be expressed as $dV = -\bar{E} \cdot d\bar{l}$, or, in scalar form, $dV = -E_l \, dl$, where E_l is the component of electric intensity in the direction dl. Dividing by dl we obtain $E_l = -dV/dl$, that is, the component of electric intensity in the direction dl is the negative space derivative of potential in this direction. In a similar manner, we may obtain components of electric intensity in the x, y, and z directions. Writing these as partial derivatives, we have

$$E_x = -\frac{\partial V}{\partial x} \qquad E_y = -\frac{\partial V}{\partial y} \qquad E_z = -\frac{\partial V}{\partial z}$$

Assigning the corresponding unit vectors and adding, we obtain the electric intensity in vector form

$$\bar{E} = -\left(\frac{\partial V}{\partial x} \bar{\imath} + \frac{\partial V}{\partial y} \bar{\jmath} + \frac{\partial V}{\partial z} \bar{k} \right) \tag{1}$$

The term in the brackets is known as the *potential gradient,* abbreviated grad V.

In vector analysis it is convenient to express mathematical relationships in terms of a *del operator*, symbolized by ∇. In rectangular coordinates, this operator is defined by

$$\nabla = \frac{\partial}{\partial x}\bar{\imath} + \frac{\partial}{\partial y}\bar{\jmath} + \frac{\partial}{\partial z}\bar{k} \tag{2}$$

If we perform the operation indicated by ∇V or "*del V*," we obtain $\nabla V = (\partial V/\partial x)\bar{\imath} + (\partial V/\partial y)\bar{\jmath} + (\partial V/\partial z)\bar{k}$, which is the same as the bracketed term in Eq. (1). Thus, ∇V is the same as grad V; hence we may express electric intensity in either one of the abbreviated forms

$$\bar{E} = -\text{ grad } V = -\nabla V \tag{3}$$

While Eq. (1) is expressed in rectangular coordinates, Eq. (3) is more general in that it may represent the electric intensity in any coordinate system. Appendix III gives the gradient in cylindrical and spherical coordinates.

A word of caution is necessary here. In a subsequent chapter,[1] we shall find that the potential V is a function of the charge distribution throughout space. Electric intensity, however, may be produced either by a distribution of charges in space or by a time-varying magnetic field. Equations (1), (3), and (2.02-5) do not take into consideration that portion of the electric intensity produced by the time-varying magnetic field; hence these equations are valid, strictly speaking, only in electrostatic fields. However, in most vacuum-tube applications, the currents involved are quite small; therefore the magnetic fields which they produce are relatively weak. Consequently, the electric intensity produced by the time variation of the magnetic field is negligible in comparison with the electric intensity resulting from the applied potentials. For this reason, Eqs. (1), (3), and (2.02-5) may be used in most vacuum-tube analyses even though the fields are not electrostatic.

2.04. Gauss's Law and Electric Flux.—Electric charges constitute a source of electric intensity or electric flux. Electric flux is customarily represented by lines starting on positive charges and terminating on negative charges. Gauss's law states that the net outward flux through any closed surface is proportional to the charge enclosed.

If a small surface element is placed perpendicular to the flux lines, the number of lines of flux per unit area is known as the *flux density*. Since the amount of flux through the surface element is dependent upon its orientation, flux density is a vector quantity which we shall designate by the symbol $\bar{D}$. In vector analysis, a differential element of area is often represented by a vector $d\bar{s}$ which is normal to the area element as shown in Fig. 6. Designating electric flux by the symbol ψ, the flux $d\psi$ through

[1] See Eqs. (15.02–3 and 11), Chap. 15.

the differential area $d\bar{s}$ is $d\psi = \bar{D} \cdot d\bar{s}$. The net outward flux through the closed surface is obtained by integrating $d\psi$ over the closed surface. In rationalized units, the net outward flux is equal to the charge enclosed; hence Gauss's law becomes

$$\psi = \oint_s \bar{D} \cdot d\bar{s} = q \tag{1}$$

The symbol $\oint_s$ denotes integration over a closed surface.
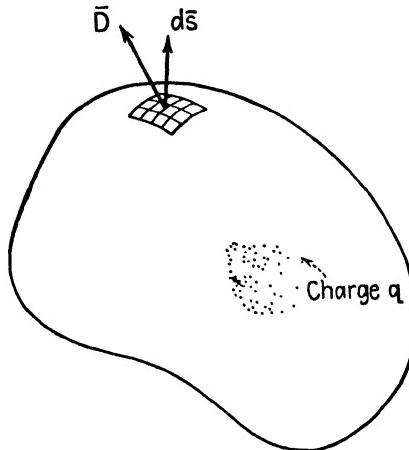
FIG. 6.—An illustration of Gauss's law.

In an isotropic dielectric, electric flux density is related to electric intensity by

$$\bar{D} = \epsilon \bar{E} \tag{2}$$

The mks unit of electric flux density is the coulomb per square meter.

2.05. Divergence and Poisson's Equation.—Gauss's law expresses a fundamental electromagnetic field relationship in integral form. We shall now derive an expression for this law in differential equation form by applying the integral form to a differential element of volume.

Consider the differential volume $d\tau$ in Fig. 7, having sides dx, dy, and dz. The *divergence of $\bar{D}$* (abbreviated div $\bar{D}$) will be defined as the net outward flux $d\psi$ through the closed surface, divided by the volume $d\tau$ or, briefly, the net outward flux per unit volume. Mathematically, the divergence of $\bar{D}$ becomes

$$\text{div } \bar{D} = \frac{d\psi}{d\tau} \tag{1}$$

Consider first the flux through surfaces a and b which are parallel to the xz plane in Fig. 7. Let $\bar{D} = D_x\bar{\imath} + D_y\bar{\jmath} + D_z\bar{k}$ be the flux density at the

center of the differential volume. Denoting outward flux density as positive, the outward components of flux density at the surfaces a and b, to a linear order of approximation, are

$$-D_y - \frac{\partial D_y}{\partial y}\left(--\frac{dy}{2}\right) \qquad \text{at surface } a$$

$$D_y + \frac{\partial D_y}{\partial y}\left(\frac{dy}{2}\right) \qquad \text{at surface } b$$
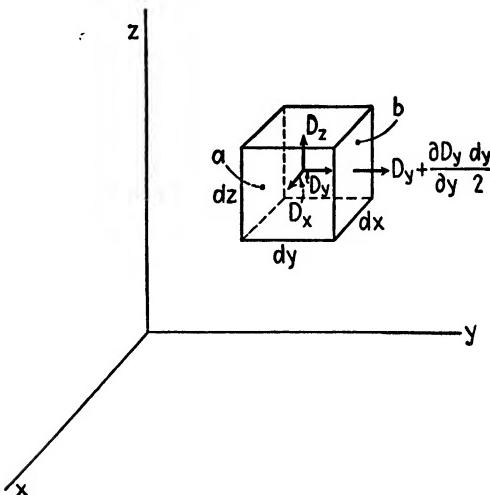
Fig. 7.—Illustration for divergence.

It is assumed that these represent the average flux densities over their respective surfaces. By multiplying the flux densities by the differential surface area $dx\,dz$ and adding to obtain the net outward flux through a and b, we have

$$d\psi_a + d\psi_b = \frac{\partial D_y}{\partial y}\,dx\,dy\,dz \qquad (2)$$

Similar expressions may be obtained for the flux through the remaining four faces. When these are added together, the net outward flux through the closed surface is

$$d\psi = \left(\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z}\right)dx\,dy\,dz \qquad (3)$$

To obtain the divergence of $\bar{D}$, substitute Eq. (3) into (1). Since the differential volume is $d\tau = dx\,dy\,dz$, the divergence becomes

$$\text{div } \bar{D} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} \qquad (4)$$

It is interesting to observe that the dot product of the del operator ∇, given by Eq. (2.03-2), and the vector $\bar{D}$ also yields Eq. (4), that is

$$\nabla \cdot \bar{D} = \operatorname{div} \bar{D} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} \tag{5}$$

Equation (5) is a mathematical expression for the divergence of $\bar{D}$. Now let us relate this to the enclosed charge, using Gauss's law.

Let q_τ be the *charge density.* The charge enclosed in the differential volume $d\tau$ is $q_\tau \, d\tau = q_\tau \, dx \, dy \, dz$. According to Gauss's law, the net outward flux $d\psi$, given by Eq. (3), must be equal to the charge enclosed in the differential element. Equating these and dividing by $d\tau$, we obtain

$$\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} = q_\tau$$

or

$$\nabla \cdot \bar{D} = q_\tau \tag{6}$$

Let us briefly consider the physical interpretation of Gauss's law and divergence. Gauss's law states that the net outward flux is equal to the charge enclosed for any closed surface. If the charge enclosed is zero, the amount of flux entering the closed surface equals the flux leaving it and the net outward flux is zero. The divergence equation is essentially Gauss's law reduced to a per-unit-volume basis, since it states that the net outward flux per unit volume is equal to the charge density. However, it should be noted that the divergence was derived for a differential volume and is therefore a point function.

We now derive another useful relationship known as Poisson's equation. First, however, let us express Eq. (6) in terms of electric intensity by substituting $\bar{E} = \bar{D}/\epsilon$, thus obtaining

$$\nabla \cdot \bar{E} = \frac{q_\tau}{\epsilon} \tag{7}$$

Now substitute $\bar{E} = -\nabla V$ from Eq. (2.03-3), to obtain

$$\nabla^2 V = -\frac{q_\tau}{\epsilon} \tag{8}$$

where $\nabla^2 = \nabla \cdot \nabla$ is the Laplacian operator. In rectangular coordinates Eq. (8) becomes

$$\nabla^2 V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = -\frac{q_\tau}{\epsilon} \tag{9}$$

Appendix III gives expressions for $\nabla \cdot \bar{A}$ and $\nabla^2 V$ in cylindrical and spherical coordinates.

Either the divergence equation or Poisson's equation may be used to determine the field distribution throughout space. The practical application of the divergence equation is limited to cases where the flux density is a function of only one coordinate. Other cases can be handled more readily by Poisson's equation. It should be noted that Gauss's law and the divergence equation are valid for either static or dynamic fields, whereas Poisson's equation is strictly valid only for electrostatic fields.

In order to illustrate the use of the foregoing relationships, consider the following example:

Example: A space charge of density $q_\tau = c\rho^{1/2}$ is contained in a cylindrical shell having an inside radius a and an outside radius b. Assume that the potential at radius a is zero. Determine the electric intensity and potential distributions inside of the cylindrical shell.

Two methods will be used, based upon Gauss's law and Poisson's equation. When $\rho < a$, the electric intensity and potential are both zero. Consider the region in which $a < \rho < b$.

A. Gauss's Law Method.—A cylinder of radius ρ, where $a < \rho < b$, and unit height would contain an amount of charge given by

$$q = \int_a^\rho q_\tau 2\pi\rho \, d\rho = \int_a^\rho 2\pi c\rho^{3/2} \, d\rho$$

$$= \frac{4\pi c}{5}(\rho^{5/2} - a^{5/2})$$

Because of symmetry, the flux density is uniform over the cylindrical surface and normal to this surface. Gauss's law may therefore be written in scalar form as follows:

$$\oint \bar{D} \cdot d\bar{s} = 2\pi\rho D_\rho = q$$

Inserting the expression for q and using $D_\rho = \epsilon E_\rho$, we obtain the electric intensity

$$E_\rho = \frac{2c}{5\epsilon\rho}(\rho^{5/2} - a^{5/2})$$

The potential is found by applying Eq. (2.02-6) to obtain

$$V = -\int E_\rho \, d\rho = -\frac{2c}{5\epsilon}\int\left(\rho^{3/2} - \frac{a^{5/2}}{\rho}\right)d\rho$$

$$= -\frac{2c}{5\epsilon}\left(\frac{2\rho^{5/2}}{5} - a^{5/2}\ln\rho + C_1\right)$$

The constant C_1 is evaluated by setting $V = 0$ when $\rho = a$. The potential then becomes

$$V = -\frac{2c}{5\epsilon}\left[\frac{2}{5}(\rho^{5/2} - a^{5/2}) - a^{5/2}\ln\frac{\rho}{a}\right]$$

B. Poisson's Equation.—From Appendix III, we obtain an expression for Poisson's equation in cylindrical coordinates which, combined with Eq. (2.05-8), gives

$$\nabla^2 V = \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial V}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 V}{\partial\phi^2} + \frac{\partial^2 V}{\partial z^2} = -\frac{q_\tau}{\epsilon}$$

Since $\partial V/\partial \phi = \partial V/\partial z = 0$, this reduces to

$$\frac{1}{\rho} \frac{d}{d\rho} \left(\rho \frac{dV}{d\rho} \right) = -\frac{c\rho^{\frac{1}{2}}}{\epsilon}$$

Multiplying by $\rho \, d\rho$ and integrating, we obtain

$$\rho \frac{dV}{d\rho} = -\frac{2c\rho^{\frac{5}{2}}}{5\epsilon} + C_2$$

This expression may be integrated a second time to obtain the potential. However, if we recall that $E_\rho = -dV/d\rho$, and that E_ρ is zero when $\rho = a$, we may evaluate the integration constant and obtain

$$E_\rho = \frac{2c}{5\epsilon\rho} (\rho^{\frac{5}{2}} - a^{\frac{5}{2}})$$

This is in agreement with the electric intensity obtained by the Gauss's law method. The potential may be evaluated by the method given in Part A.

The above methods can be used to evaluate the electric intensity and potential outside of the cylinder b. However, in using Poisson's equation, it should be noted that the space-charge density q_r is zero in regions where $\rho > b$. These solutions yield one integration constant in the electric intensity equation and two in the potential equation. The constants may be evaluated in terms of the previously determined values of electric intensity and potential at radius b.

2.06. Motion of Charges in Electric Fields.—Having considered the laws determining the electric-field distribution in space, we now turn to a consideration of the motion of charges in electric fields.

The force on a charge q in an electric field is given by

$$\bar{f} = q\bar{E} \tag{1}$$

If the charge is free to move in space, the acceleration of the charge is given by Newton's second law of motion,

$$q\bar{E} = m \frac{d^2\bar{l}}{dt^2} \tag{2}$$

where m is the mass in kilograms and $d^2\bar{l}/dt^2$ is the acceleration in meters per second per second.

Equation (2) states that the acceleration is proportional to electric intensity. This equation may be resolved into components in any coordinate system. In rectangular coordinates this becomes

$$\frac{d^2x}{dt^2} = \frac{qE_x}{m}$$

$$\frac{d^2y}{dt^2} = \frac{qE_y}{m}$$

$$\frac{d^2z}{dt^2} = \frac{qE_z}{m} \tag{3}$$

The electric intensity may be determined as a function of space and time coordinates using the previously derived relationships. Since these are first-order differential equations in velocity and second-order equations in displacement, the velocity equation contains one integration constant and the displacement equation has two integration constants. These constants can be evaluated if the position and velocity of the charge are known at any instant of time.

The equations for the velocity and energy of an electron may also be expressed in terms of potential. In Fig. 8, the work done by the field in moving a charge q from point a to point b is equal to the line integral of force times distance. The force is $q\bar{E}$, hence the work becomes $w = \displaystyle\int_a^b q\bar{E}\cdot d\bar{l}$. Invoking the law of conservation of energy, we find that the work done by the field in moving the charge from a to b is equal to the increase in kinetic energy of the charge. The kinetic energy of a particle is $\frac{1}{2}mv^2$, where v is its velocity, hence
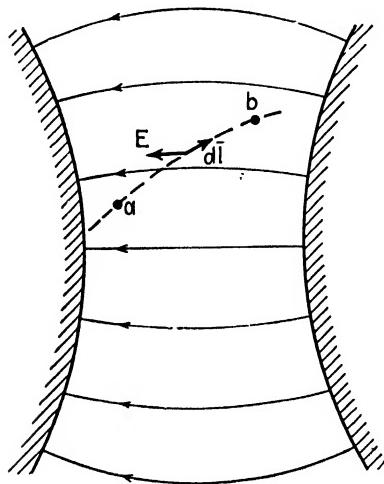
FIG. 8.—Motion of a charge in an electric field.

$$q\int_a^b \bar{E}\cdot d\bar{l} = \tfrac{1}{2}m(v_b^2 - v_a^2) \tag{4}$$

For electrostatic fields we may substitute Eq. (2.02-5) for the integral on the left-hand side of Eq. (4). Equation (4) then becomes

$$-q(V_b - V_a) = \tfrac{1}{2}m(v_b^2 - v_a^2) \tag{5}$$

Solving for v_b, we obtain

$$v_b = \sqrt{v_a^2 - \frac{2q}{m}(V_b - V_a)} \tag{6}$$

If the potential and velocity are both zero at a, this reduces to

$$v_b = \sqrt{-\frac{2qV_b}{m}} \tag{7}$$

The negative sign under the radical of Eq. (7) signifies that a positive charge gains in velocity as it moves toward a negative potential.

Equations (5) to (7) are, strictly speaking, valid only for electrostatic fields. They may, however, be used for dynamic fields provided that there

is no appreciable time variation of the field during the flight of the electron.

In vacuum-tube analysis we are primarily concerned with electron motion, in which case we have

$$q = -e = -1.602 \times 10^{-19} \text{ coulombs}$$

$$m = 9.107 \times 10^{-31} \text{ kilogram}$$

$$\frac{e}{m} = 1.759 \times 10^{11} \text{ coulombs per kilogram}$$

If an electron moves at a high velocity, its apparent mass is greater than its stationary mass. The apparent mass is known as the relativistic mass and is related to the stationary mass m_0 as follows:

$$m = \frac{m_0}{\sqrt{1 - (v/v_c)^2}} \qquad (8)$$

In this equation v is the velocity of the electron in meters per second and $v_c = 3 \times 10^8$ meters per second is the velocity of light.

For a 1 per cent increase in mass the velocity must equal 14 per cent of the velocity of light. To attain this velocity an electron must start from rest and be accelerated through a potential difference of 5,100 volts. Since most vacuum tubes

Fig. 9.—Temperature-limited diode with d-c and a-c applied potentials.

operate with potential differences which are less than 5,100 volts, the relativistic correction can usually be ignored.

If the relativistic correction must be taken into account, the more fundamental force equation which states that force is equal to the time rate of change of momentum, or

$$\bar{f} = \frac{d(m\bar{v})}{dt} \qquad (9)$$

must be used in place of $\bar{f} = m\bar{a}$.

2.07. Electron Motion in Time-varying Fields.—In tubes operating at microwave frequencies, the time of transit of an electron between two electrodes in the tube may be relatively large in comparison with the period of electrical oscillation. It is interesting to study the motion of electrons through the fields when large transit times are involved.

As an example, let us investigate the motion of electrons in the time-varying field of a parallel-plane diode in which the current is temperature limited. In the diode of Fig. 9, the instantaneous potential difference be-

tween cathode and plate is assumed to have a direct component V_0 and an alternating component $V_1 \sin \omega t$, thus
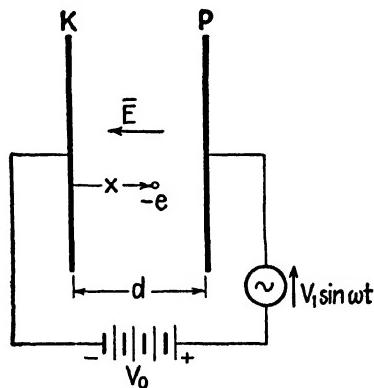
$$V = V_0 + V_1 \sin \omega t \tag{1}$$

Since the space-charge density is negligible, the electric intensity in the diode space is given by

$$E_x = -\frac{1}{d}(V_0 + V_1 \sin \omega t) \tag{2}$$

Substituting the electron charge $q = -e$ and Eq. (2) in the first of (2.06-3), we have

$$\frac{d^2x}{dt^2} = \frac{e}{md}(V_0 + V_1 \sin \omega t) \tag{3}$$

Two successive integrations of Eq. (3) yield the electron velocity and displacement as a function of time. Assuming that the electron leaves the cathode ($x = 0$) at time t_0 and with velocity v_0, we obtain

$$\frac{dx}{dt} = \frac{e}{md}\left[V_0(t - t_0) - \frac{V_1}{\omega}(\cos \omega t - \cos \omega t_0)\right] + v_0 \tag{4}$$

$$x = \frac{e}{md}\left[V_0\frac{(t - t_0)^2}{2} - \frac{V_1}{\omega^2}(\sin \omega t - \sin \omega t_0) + \frac{V_1}{\omega}(t - t_0)\cos \omega t_0\right]$$
$$+ v_0(t - t_0) \tag{5}$$

The time $t - t_0$ is the *electron transit time*, or the time required for the electron to travel from the cathode to a point distant x from the cathode. This electron transit time will be designated by T. The *total transit time* T_1 represents the time required for the electron to travel the entire distance from cathode to anode. We now let $\phi = \omega t_0$ be the phase angle of alternating potential at the instant of electron departure from the cathode. We therefore have

$$T = t - t_0 \tag{6}$$

$$\phi = \omega t_0 \tag{7}$$

Now write Eq. (5) in terms of T and t_0, using Eq. (6). Then substitute Eq. (7) and divide both sides of the equation by a factor k, which is defined below, to obtain

$$\frac{x}{k} = \frac{(\omega T)^2}{2} + \frac{V_1}{V_0}[(\omega T - \sin \omega T)\cos \phi + (1 - \cos \omega T)\sin \phi] + v_0\frac{(\omega T)}{\omega k} \tag{8}$$

$$= A + \frac{V_1}{V_0}B + C \tag{9}$$

where $k = \dfrac{eV_0}{\omega^2 md} = 1.76 \times 10^{11} \dfrac{V_0}{\omega^2 d}$ (mks units)

$$A = \frac{(\omega T)^2}{2}$$

$$B = (\omega T - \sin \omega T) \cos \phi + (1 - \cos \omega T) \sin \phi \qquad (10)$$

$$C = v_0 \frac{(\omega T)}{\omega k}$$

The electron transit time in a pure d-c field with zero initial velocity is obtained by setting $V_1 = 0$ and $v_0 = 0$ in Eq. (8) and solving for time. Designating the total d-c transit time as T_0, we have

$$T_0 = d\sqrt{\frac{2m}{eV_0}} = 3.37 \times 10^{-6} \frac{d}{\sqrt{V_0}} \qquad (11)$$

The quantity ωT is the *transit angle*, representing the number of radians through which the alternating potential varies during the transit time T. The quantity A in Eq. (9) represents the value which the displacement parameter x/k would have if there were no alternating potential difference and the initial electron velocity were zero. Thus, we may consider A as being the d-c component of the displacement parameter x/k. Graphs of A as a function of transit angle ωT, as expressed in Eq. (10), are given in Fig. 10a for small transit angles and in Fig. 10b for large angles.

The term $(V_1/V_0)B$ in Eq. (9) is the component of x/k resulting from the alternating field. Values of B are plotted in Figs. 11a and 11b as functions of transit angle for various values of departing phase angle ϕ. The curves in Fig. 11 show that the a-c component of the field alternately accelerates and retards the electron in its flight. For relatively large transit angles, the a-c component of displacement is a maximum for $\phi = 0$ degrees, *i.e.*, when the electron leaves the cathode just as the alternating potential is changing from deceleration to acceleration.

The quantity C in Eq. (9) represents the component of x/k due to the initial velocity of the electron. The value of C is zero if the initial velocity is zero.

The electron displacement for a given transit angle may be readily obtained by the use of Figs. 10 and 11, and Eq. (9). If the transit angle ωT and departing phase angle ϕ are given, the values of A and B may be obtained directly from the curves. The value of C may be computed from Eq. (10). Substitution of these values in Eq. (9) yields the value of x/k and this may be used to determine the value of electron displacement x during the time T.
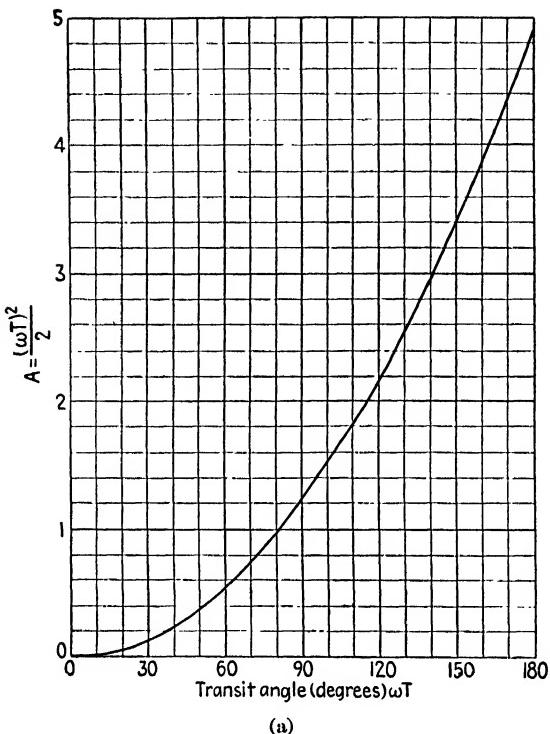
(a)
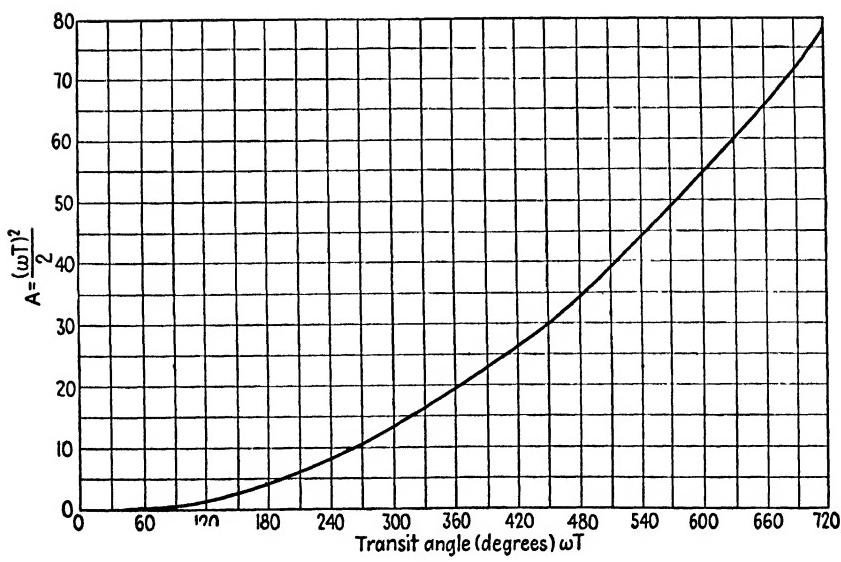
(b)

Fig. 10.—D-c component of x/k as a function of transit angle.

(a)

(b)
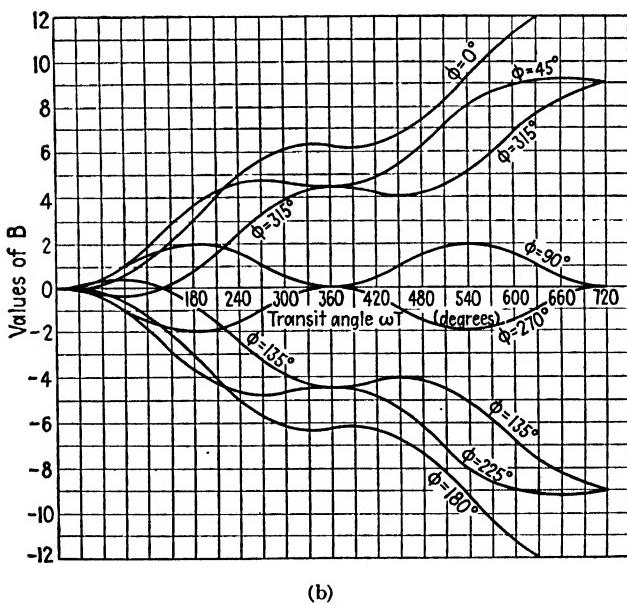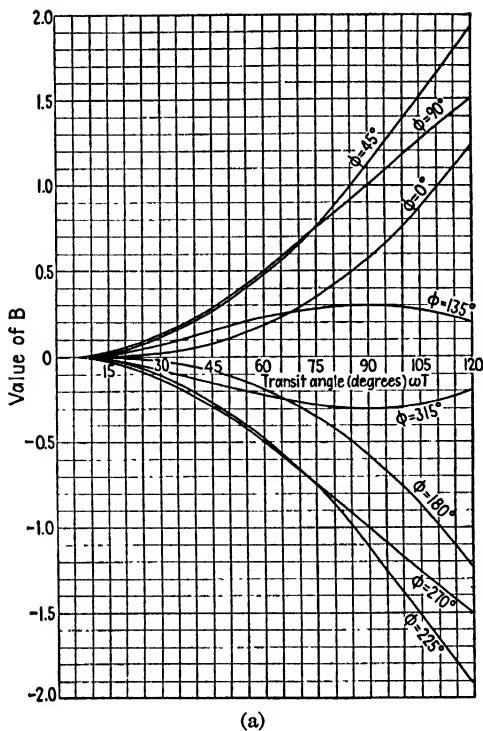
FIG. 11.—A-c component of x/k as a function of transit angle.

The reverse process, *i.e.*, finding the transit angle corresponding to a given electron displacement, requires a trial-and-error procedure. The value of x/k is first computed and the d-c transit angle corresponding to this value of x/k is obtained from Fig. 10. This transit angle is used as a first approximation, and the corresponding values of B and C are determined from Fig. 11 and Eq. (10). The value of x/k computed from Eq. (9) using the first approximations for A, B, and C will, in general, not agree with the correct value computed previously. It will therefore be necessary to assume new values of ωT in the vicinity of the first approximation until one is found which yields A, B, and C values satisfying Eq. (9).

Example. To illustrate the solution of a typical problem, let us determine the transit angle and total transit time required for an electron to travel from cathode to anode in a temperature-limited parallel-plane diode having the following values:

$$d = 0.5 \text{ cm} \qquad \phi = 0°$$

$$V_0 = 1,000 \text{ volts} \qquad v_0 = 0$$

$$V_1 = 800 \text{ volts} \qquad f = 2 \times 10^9 \text{ cycles per sec}$$

First obtain the values

$$\frac{V_1}{V_0} = 0.8 \qquad \text{and} \qquad k = \frac{1.76 \times 10^{11} V_0}{\omega^2 d} = 2.224 \times 10^{-4}$$

The electron travels a distance $x = d = 5 \times 10^{-3}$ m. Thus, we have $x/k = 22.45$.

If we consider only the d-c potential, Fig. 10b shows that the transit angle corresponding to $A = 22.45$ is $\omega T = 380$ degrees. This is the first approximation to the value of ωT. Figure 11b shows that the value of B corresponding to this approximate transit angle is $B = 6.3$, and thus $(V_1/V_0)B = 5.04$. It is apparent, therefore, that the value of x/k using these values of A and B is too high by approximately the amount of $(V_1/V_0)B$.

As a second approximation, therefore, try $A = 22.45 - 5.04 = 17.41$. Figure 10b yields the second approximation $\omega T = 345°$ and Fig. 11b gives the corresponding value $B = 6.3$, or $(V_1/V_0)B = 5.04$. Substituting these in Eq. (9) we obtain $x/k = 22.45$ which is the correct value. Thus, the transit angle is $\omega T = 345°$ or 6.02 radians and the total transit time is $T = 6.02/\omega = 4.78 \times 10^{-10}$ sec.

If the field has no d-c component, we have $V_0 = 0$. In order to evaluate Eq. (8), it is first necessary to multiply both sides by V_0, yielding

$$x = V_1 k'B + v_0 T \tag{12}$$

where $k' = \dfrac{e}{\omega^2 md} = \dfrac{1.76 \times 10^{11}}{\omega^2 d}$ (mks units)

If the electron enters the a-c field with a high initial velocity, the transit time in traveling a distance x may be approximated by $T = x/v_0$. The corresponding value of ωT is then computed and the value of B is obtained from Fig. 11. The amount of error involved in the original assumption can be obtained by inserting these values into Eq. (12). Other values of T are then assumed until one is found which does satisfy Eq. (12).

PROBLEMS

1. Two vectors $\bar{A}$ and $\bar{B}$ are given by

$$\bar{A} = 2\bar{\imath} + 8\bar{\jmath} + 6\bar{k}$$

$$\bar{B} = 2\bar{\imath} + 1\bar{\jmath} - 2\bar{k}$$

    (a) Compute the vectors represented by $\bar{A} + \bar{B}$ and $\bar{A} - \bar{B}$

    (b) Compute the dot and cross products $\bar{A} \cdot \bar{B}$ and $\bar{A} \times \bar{B}$

    (c) The angles between a vector and the x, y, and z axis, respectively, may be obtained from the direction cosines. The direction cosines of the vector $\bar{A}$ are $l = \cos \theta_x = A_x/A$, $m = \cos \theta_y = A_y/A$, $n = \cos \theta_z = A_z/A$. Compute the direction cosines for the vectors $\bar{A}$ and $\bar{B}$ given above.

    (d) Show that the cosine of the angle between two vectors is given by $\cos \theta_{AB} = l_A l_B + m_A m_B + n_A n_B$. Find the angle θ_{AB} for the vectors given above.

2. The vectors $\bar{A}$, $\bar{B}$ and $\bar{C}$ are given by

$$\bar{A} = 3\bar{\imath} - 1\bar{\jmath} + 2\bar{k}$$

$$\bar{B} = 1\bar{\imath} + 2\bar{\jmath} - 1\bar{k}$$

$$\bar{C} = 2\bar{\imath} - 2\bar{\jmath} - 2\bar{k}$$

Find the values of

    (a) $\bar{A} \cdot (\bar{B} \times \bar{C})$ and

    (b) $\bar{A} \times (\bar{B} \times \bar{C})$.

3. Derive equations in rationalized mks units for the electric intensity and potential distribution as functions of distance from

    (a) A point charge

    (b) A uniformly charged infinitely long line having a charge of q_l coulombs per meter of length

    (c) An infinite plane having a charge density of q_r coulombs per square meter.

    Note: The variation of field intensity and potential with respect to distance for point, line, and plane sources is the same for electric fields, magnetic fields, gravitational fields, light fields, and many other types of fields.

4. An electrostatic field has an intensity distribution in the xy plane given by $\bar{E} = (C_1/x)\bar{\imath} + (C_2/y)\bar{\jmath}$. Evaluate the line integral $\int \bar{E} \cdot d\bar{l}$ over any path from $(x = 1, y = 2)$ to $(x = 3, y = 3)$. Hint: The integration is simplified by taking a path parallel to the x axis from $(1, 2)$ to $(3, 2)$, then parallel to the y axis from $(3, 2)$ to $(3, 3)$.

5. (a) Derive an expression for the difference of potential between the conductors of a coaxial transmission line. Assume that the inner conductor has a charge of q_l coulombs per meter of length and the outer conductor has a charge of $-q_l$ coulombs per meter of length.

    (b) The capacitance per unit length is the charge per unit length divided by the difference of potential between conductors. Obtain an expression for the capacitance per unit.

6. A point charge q is placed at the point $(x = -2, y = +2, z = 0)$.

    (a) Write an expression for the electric intensity at any point in space.

    (b) Evaluate the difference of potential between the points $(0, 0, 0)$ and $(2, 4, 0)$ by integrating along the curve $y = x^2$.

7. A point charge is located at the origin of a rectangular coordinate system. Consider a circular plane of unit radius which is oriented in a position normal to the x axis, with center at $(x = 1, y = 0, z = 0)$.

  (a) Write the equation for the potential on the surface of the circular plane.

  (b) Using the gradient relationships, determine the tangential and normal components of electric intensity at the surface.

  (c) Find the surface integral $\int \bar{D} \cdot d\bar{s}$ over the given area.

8. Derive the equations for grad V in cylindrical and spherical coordinates.

9. Derive the equation for $\nabla \cdot \bar{D}$ in cylindrical coordinates (see Appendix III).

10. An insulating cylinder of radius a contains a uniform charge of density q_r.

  (a) Using Gauss's law, derive expressions for the electric intensity and potential at points (1) inside the cylinder and (2) outside the cylinder.

  (b) Derive these relationships using either Poisson's equation or the relationship $\nabla \cdot \bar{D} = q_r$ in cylindrical coordinates.

11. In a spherical coordinate system the space-charge density is $q_r = C(r)^{1/2}$.

  (a) Using Gauss's law, derive expressions for the electric intensity and potential as functions of r.

  (b) Repeat part (a) starting with either Poisson's equation or the relationship $\nabla \cdot \bar{D} = q_r$ expressed in spherical coordinates.

  (c) Evaluate $\oint \bar{D} \cdot d\bar{s}$ over the surface of a sphere of radius a with center at the origin.

  (d) Compute $\int \nabla \cdot \bar{D} \, d\tau$. Show that $\oint \bar{D} \cdot d\bar{s} = \int_\tau \nabla \cdot \bar{D} \, d\tau$. This is a theorem of vector analysis known as the *divergence theorem*.

12. A cylindrical diode has a potential difference of V_b volts between cathode and anode. The radii of the cathode and anode are a and b, respectively. Assuming negligible space-charge density, derive equations for the acceleration, velocity, and displacement of an electron which is traveling radially outward from the cathode.

  Note: In this problem, an integral of the form $\int dx/(\ln x)^{1/2}$ is encountered. To evaluate this integral, substitute $x = e^y$ and integrate by series methods.

13. A parallel-plane diode with space-charge-limited emission has a space-charge density given by $q_r = -C/(x)^{2/3}$.

  (a) Derive expressions for the electric intensity and potential distribution in the diode.

  (b) Derive equations for the acceleration, velocity, and displacement of an electron. Compare the total electron transit time in the space-charge-limited diode with Eq. (2.07-11) for the diode with temperature-limited emission.

14. In a klystron oscillator, an electron starts from rest and is accelerated through a d-c potential difference of 400 volts. It then passes through the region between two parallel-plane grids in a direction normal to the grids. The grids are 0.2 cm apart and have an a-c potential difference of 350 volts (peak value). The frequency is 3,000 megacycles per sec. It is assumed that there is no d-c field between the grids. Determine the transit time and transit angle for an electron which enters the grid region as the electric field is passing through zero, changing from acceleration to deceleration.

CHAPTER 3

CURRENT, POWER, AND ENERGY RELATIONSHIPS

Vacuum-tube oscillators and amplifiers are essentially devices for converting d-c energy into a-c energy. Such devices contain two functionally different but interdependent parts: (1) the vacuum tube with its stream of electrons moving under the influence of electric and magnetic fields and (2) the external circuit containing, among other things, the sources of potential and the load impedance.

A complete analysis of such a vacuum-tube system would require an analysis of the vacuum tube from the point of view of electron dynamics and an analysis of the external circuit from a circuit viewpoint. These two solutions are interdependent since the electronic effects influence the potentials in the external circuit and the potentials, in turn, determine the fields in which the electrons move. Since a rigorous treatment of such a system involves considerable mathematical difficulty, it is customary to use either of two simplified methods of approach. In the conventional method, the tube is represented by an equivalent generator and equivalent internal impedances. This equivalent circuit is then joined to the external circuit and the analysis proceeds as an electric-circuit problem. Although this method greatly simplifies the analysis of vacuum-tube circuits, it loses sight of the fundamental electronic phenomena taking place inside of the vacuum tube.

The second method deals largely with the electronic phenomena within the tube. In this method, certain arbitrary direct and alternating potentials are assumed to exist at the tube terminals without inquiring as to what external conditions are required to produce the assumed potentials. The behavior of the electrons and their electrical effects are then expressed in terms of the assumed potentials. An equivalent circuit for the vacuum tube may also be obtained by this method of analysis. In general, however, this equivalent circuit will differ from the equivalent circuit of the preceding method and will represent more accurately the conditions existing at frequencies where electron transit-time effects are significant.

In this chapter, we shall consider the concepts of current, power, and energy from a fundamental electronic point of view, using the electronic method of approach.

3.01. Convection and Conduction Current.—The motion of electric charges constitutes an electric current. Charges moving in space consti-

tute a *convection current*, whereas the motion of charges in a conductor constitutes a *conduction current*. In either case, the convection or conduction current density J_c is equal to the product of charge density times velocity, or

$$J_c = q_r \bar{v} \tag{1}$$

where q_r is the charge density and $\bar{v}$ is its velocity.

In a conducting medium the moving charges experience a frictional resistance. In most conductors,[1] the average velocity of the charges is proportional to the electric intensity $\bar{E}$ and Eq. (1) may therefore be written

$$J_c = \sigma \bar{E} \tag{2}$$

where σ is a property of the medium known as the *conductivity*. In mks units, J_c is in amperes per square meter and σ is in mhos per meter. Values of σ for various conducting mediums are given in Appendix II.

FIG. 1.—Conduction current in a conductor.

Equation (2) is Ohm's law expressed in terms of current density and electric intensity. To relate this equation to the more familiar form of Ohm's law, consider a direct current flowing in the homogeneous cylindrical conductor of Fig. 1. Assume that the current density is uniform over the cross section of the conductor. The total conduction current i_c is then the product of current density times area, or $i_c = J_c A$. The electric intensity is uniform throughout the conductor, hence the potential drop over a length of conductor l is $V_R = El$. Substituting J_c and E from these two relationships into Eq. (2), written in scalar form, we obtain Ohm's law

$$i_c = \frac{\sigma A}{l} V_R = \frac{V_R}{R} \tag{3}$$

where $R = l/\sigma A$ is the electrical resistance of the given length of conductor.

3.02. Continuity of Current—Displacement Current.—Kirchhoff formulated an important law of continuity of conduction current in closed circuits. This law states that in an electric circuit the current flowing to a point is equal to the current flowing away from the point.

Maxwell introduced the concept of displacement current and thereby generalized the law of continuity of current. Consider, for example, an a-c circuit containing a condenser. If the condenser is charging or discharging, a conduction current flows in the metallic circuit. Since the

---

[1] Joos, GEORGE, "Theoretical Physics," p. 425, G. E. Stechert & Company, New York, 1934.

charges do not flow through the dielectric of the condenser, the conduction current is discontinuous at the condenser plates. Hence, if we take the restricted viewpoint that current consists only of the flow of charges, it follows that current is not always continuous. Maxwell showed that if the time variation of the electric field is treated as a displacement current, then current is always continuous; *i.e.*, current always flows in closed paths. In the example cited above, the displacement current between the condenser plates resulting from the time variation of the electric field is exactly equal to the conduction current in the external circuit. It was this reasoning that made it possible for Maxwell to predict the propagation of electromagnetic waves through space.

FIG. 2.—Continuity of current in a discharging condenser circuit.

In order to obtain an expression for the displacement current, consider the condenser circuit of Fig. 2 during the time that the condenser is discharging.[1] The positive plate of the condenser is assumed to be totally enclosed by an imaginary hemispherical shell. A positive sign will be used to denote current flowing out of the hemispherical surface. The principle of conservation of electricity states that the conduction current flowing out through the hemispherical surface must equal the time rate of decrease of charge on the condenser plate. Thus, if q is the charge on the positive plate, the conduction current flowing in the external circuit is

$$i_c = -\frac{dq}{dt} \tag{1}$$

Gauss's law relates the electric flux through the hemispherical surface to the charge enclosed,

$$\oint_s \overline{D} \cdot d\overline{s} = q \tag{2.04-1}$$

By inserting q from this expression into Eq. (1) and interchanging the order of differentiation and integration, we obtain

$$i_c + \oint_s \frac{\partial \overline{D}}{\partial t} \cdot d\overline{s} = 0 \tag{2}$$

[1] FRANK, N. H., "Introduction to Electricity and Optics," Chap. 8, McGraw-Hill Book Company, Inc., New York, 1940.

We may now express the conduction current as the surface integral of the conduction-current density, thus, $i_c = \oint_s \bar{J}_c \cdot d\bar{s}$. Inserting this into Eq. (2) and combining the terms under one integral, we obtain

$$\oint_s \left(\bar{J}_c + \frac{\partial \bar{D}}{\partial t} \right) \cdot d\bar{s} = 0 \tag{3}$$

Equation (3) is a statement of the law of continuity of current. The terms $\oint_s \bar{J}_c \cdot d\bar{s}$ and $\oint_s (\partial \bar{D}/\partial t) \cdot d\bar{s}$ represent, respectively, the conduction and displacement currents through the hemispherical surface. Equation (3) states that the net current out through the hemispherical surface is equal to zero or that the current entering this surface is equal to the current leaving it. Designating the displacement current density by $\bar{J}_d$, we have

$$J_d = \frac{\partial \bar{D}}{\partial t} = \epsilon \frac{\partial \bar{E}}{\partial t} \tag{4}$$

In Fig. 2, the conduction current is in the conductor while the displacement current is a result of the time variation of the electric field between the condenser plates. In the more general case, displacement current and conduction (or convection) current may occur coincidentally. The current density is then

$$\bar{J} = \bar{J}_c + \epsilon \frac{\partial \bar{E}}{\partial t} \tag{5}$$

In Sec. 2.05, the divergence of a vector was defined as the net outward flux per unit volume. According to the law of continuity of current, the net outward current through a closed surface is always zero; hence we have

$$\nabla \cdot \bar{J} = 0 \tag{6}$$

By inserting Eq. (5) into (6), with $\bar{J}_d = \partial \bar{D}/\partial t$, we obtain, after interchanging the order of differentiation, $\nabla \cdot \bar{J}_c + \partial(\nabla \cdot \bar{D})/\partial t = 0$. Further substitution of $\nabla \cdot \bar{D} = q_\tau$ from Eq. (2.05-6) gives

$$\nabla \cdot \bar{J}_c = -\frac{\partial q_\tau}{\partial t} \tag{7}$$

Equation (7) is the *equation of continuity*. It is similar to Eq. (1) except that (7) is stated in differential equation form while Eq. (1) is in integral form.

3.03. Current Resulting from the Motion of Charges.—Consider the current resulting from the motion of a single charge between the parallel planes of the diode of Fig. 3. In Fig. 3a, the electron is passing through

plane A; hence the current through this plane is a convection current at the instant shown. In order to satisfy the principle of continuity of current, there must be equal currents flowing simultaneously through all planes parallel to plane A in the interelectrode space and in the external circuit. The electron has an electric field associated with it. Therefore, if the electron is in motion, there will be a time-varying electric field at plane B (and at all other planes parallel to plane A in the diode space). This varying electric field constitutes a displacement current. The displacement current through all planes parallel to plane A, in Fig. 3a, is exactly equal to the convection current through plane A.

Now consider the current flowing in the external circuit. The motion of the electron in the diode space tends to induce a potential difference between the planes of the diode. Hence charges flow in the external circuit in such a manner as to tend to maintain the diode planes at the same potential. This flow of charges in the external circuit constitutes an *induced conduction current*, or briefly an *induced current*.

Our complete picture of the current resulting from the motion of the single electron in Fig. 3a therefore includes a convection current through plane A, a displacement current through all planes parallel to plane A in the diode space, and an induced current in the external circuit, all flowing simultaneously and satisfying the law of continuity of current.

The two planes of the diode form a condenser. If an alternating potential difference is applied between the planes, there will be an additional displacement current in the diode space owing to the capacitance between the planes, and an equal current flowing in the external circuit. This current will be referred to as the *capacitive current*. The capacitive current is independent of the flow of charges in the diode space.

Before leaving the subject of current, let us make one further interesting observation. We have thus far considered conduction and convection currents as being quite different from displacement current. The first two were attributed to the motion of discrete charged particles, whereas displacement current was assumed to result from a time variation of the electric field. However, if we focus our attention entirely upon the field of the charge and set aside our concept of the charge as being a small particle, then we have a unified point of view in which all current becomes
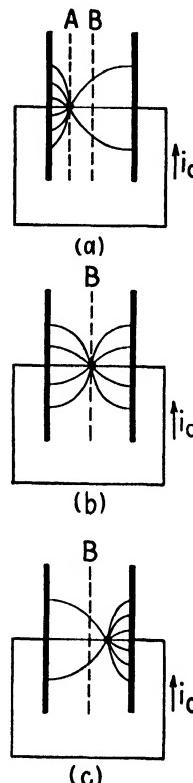
Fig. 3.—Current resulting from the motion of a single electron.

displacement current and we no longer need the separate concepts of convection and conduction current. From this viewpoint, the charges constitute the source of the electric field and the current is due entirely to the time variation of the electric field rather than to the motion of the charged particles *per se*. There is considerable justification for this point of view.

3.04. Power and Energy Relationships for a Single Electron.[1-3]—Let us briefly review the mechanics of a particle in motion. If a particle experiences a vector force $\bar{f}$, the work done on the particle in traveling a differential distance dl is

$$dw = \bar{f} \cdot d\bar{l} \tag{1}$$

or the total work done in traveling from a to b is

$$w = \int_a^b \bar{f} \cdot d\bar{l} \tag{2}$$

If the particle experiences no other force, the work done on the particle is equal to the gain in its kinetic energy. To show this, we write Newton's second law of motion

$$\bar{f} = m \frac{d^2\bar{l}}{dt^2} \tag{3}$$

where $d^2\bar{l}/dt^2$ is the acceleration of the particle. Now multiply the left-hand side of Eq. (3) by $d\bar{l}$ and the right-hand side by $(d\bar{l}/dt)\,dt$ and integrate, yielding

$$\int \bar{f} \cdot d\bar{l} = \int m \frac{d^2\bar{l}}{dt^2} \cdot \frac{d\bar{l}}{dt} dt$$

This may be written [4]

$$\int \bar{f} \cdot d\bar{l} = \frac{1}{2} m \int d \left(\frac{dl}{dt} \right)^2 \tag{4}$$

Assume that the particle moves from a to b and that the velocities at these two points are v_a and v_b. The limits on the left-hand side of Eq. (4) are

[1] JEN, C. K., On the Induced Current and Energy Balance in Electronics, *Proc. I.R.E.*, vol. 29, pp. 345–349; June, 1941.

[2] JEN, C. K., On the Energy Equation in Electronics at Ultra-High Frequencies, *Proc. I.R.E.*, vol. 29, pp. 464–466; August, 1941.

[3] GABOR, D., Energy Conversion in Electronic Devices, *J.I.E.E.*, vol. 91, pp. 128–145; September, 1944.

[4] To prove this, write $\frac{1}{2}d(dl/dt)^2$ from the right-hand side of Eq. (4) in the form $\frac{1}{2} d \left(\frac{dl}{dt} \right)^2 = \frac{1}{2} \frac{d}{dt} \left(\frac{d\bar{l}}{dt} \cdot \frac{d\bar{l}}{dt} \right) dt$. Differentiating the right-hand side of this expression as a product, we obtain $\frac{1}{2} d \left(\frac{dl}{dt} \right)^2 = \frac{d^2\bar{l}}{dt^2} \cdot \frac{d\bar{l}}{dt} dt$. This is the substitution made in obtaining Eq. (4).

a and b, while those on the right-hand side are v_a and v_b. Hence, Eq. (4) becomes

$$\int_a^b \bar{f} \cdot d\bar{l} = \tfrac{1}{2}m(v_b^2 - v_a^2) \tag{5}$$

This result shows that the work done on the particle is equal to its gain in kinetic energy.

Now consider the power relationships. Power is the time rate of change of energy. Returning to Eq. (1), if the work dw is done in time dt, the power received by the particle is

$$p = \frac{dw}{dt} = \bar{f} \cdot \frac{d\bar{l}}{dt} \tag{6}$$

i.e., the power is equal to force times velocity.
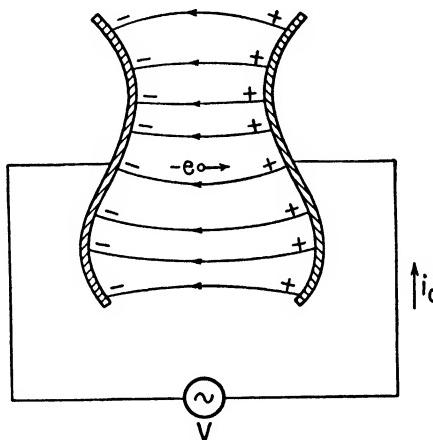
FIG. 4.—Electron motion in a diode of arbitrary geometry.

The total work done on the particle may also be expressed as the time integral of power. Combining this with Eq. (2), we have

$$\int_a^b \bar{f} \cdot d\bar{l} = \int_{t_1}^{t_2} p\, dt \tag{7}$$

where the particle is assumed to leave point a at time t_1 and arrive at b at time t_2.

Let us now apply these relationships to the case of an electron moving in the electric field of a diode of arbitrary geometry. Assume that the potential difference between the electrodes is V, where V is a function of time. The resulting electric intensity $\bar{E}$ is a function of space coordinates and time.

An electron in the diode space experiences a force $\bar{f} = -e\bar{E}$. Substitution of this force into Eq. (6) gives the power transfer from the field to the electron, $p = -e\bar{E}\cdot d\bar{l}/dt$. We have found that the moving electron induces a current i_c in the external circuit. The source of potential therefore supplies an amount of power equal to $p = i_c V$ to the field,[1] this being exactly equal to the power transferred from the field to the electron. Equating the two power expressions and writing the velocity as $\bar{v}$, we obtain

$$p = -e\bar{v}\cdot\bar{E} = i_c V \tag{8}$$

The induced current obtained from this expression is

$$i_c = -\frac{e\bar{v}\cdot\bar{E}}{V} \tag{9}$$

The increase in kinetic energy of the charge as it moves from a to b is obtained by inserting $\bar{f} = -e\bar{E}$ into Eq. (5), giving

$$w = \tfrac{1}{2}m(v_b^2 - v_a^2) = -e\int_a^b \bar{E}\cdot d\bar{l} \tag{10}$$

The line integral in Eq. (10) is similar to the expression for the potential difference between a and b as given by Eq. (2.02-5) and we might be tempted to make this substitution. However, we must bear in mind that the potential and the electric intensity may both vary appreciably with respect to time while the electron is in flight along the path from a to b. Hence, the line integral in Eq. (10) is not the potential difference at any instant of time and we are not justified in considering it as a potential difference unless the field is substantially constant during the time of transit of the electron.

The fundamental concepts of power and energy transfer between the potential source and a moving electron are embodied in Eqs. (8) and (10). In this process, the electron is accelerated by the electric field; hence there is a transfer of energy from the field to the electron. The motion of the electron in the interelectrode space induces a current in the external circuit. The induced current flowing in the external circuit transfers an equal amount of energy from the source of potential to the field to make up for the energy transfer from the field to the electron. Thus, in effect, energy is transferred from the source of potential to the moving electron. A positive sign denotes energy transfer from the source of potential to the elec-

---

[1] This is the power supplied to the field due to the motion of the electron in the field. If the potential is varying with respect to time, the potential source may supply additional capacitive power to the system. However, we are primarily interested in the power and energy transfer between the potential source and the electron and hence will not consider the capacitive power.

tron, whereas a negative sign denotes energy transfer from the electron to the source of potential.

One important concept must not be overlooked. The energy transfer occurs *while the electron is in flight*. The moment that the electron collides with an electrode, its remaining energy, as given by Eq. (10), is dissipated.

For the special case of a parallel-plane diode, we have $E = -V/d$ and Eqs. (8) and (9) become

$$p = \frac{evV}{d} = i_c V \qquad (11)$$

$$i_c = \frac{ev}{d} \qquad (12)$$

The velocity of the electron may be evaluated by methods described in the preceding chapter. It is significant to note that the current is dependent not only upon the charge and its velocity but also upon the distance between the diode planes.

3.05. Single Electron in Superimposed D-C and A-C Fields.—In most vacuum-tube applications, the electrode potentials have d-c and a-c components; hence the electrons

FIG. 5.—Diode with d-c and a-c applied potentials.

move through combined d-c and a-c electric fields. Under proper conditions, the electrons may take energy from the source of d-c potential and transfer a portion of this to the source of a-c potential.

Consider the diode of Fig. 5, which is assumed to have negligible space charge. The instantaneous potential difference between plate and cathode is

$$V = V_0 + V_1 \sin \omega t \qquad (1)$$

where V_0 is the d-c potential and $V_1 \sin \omega t$ is the a-c potential. The electric intensity in the interelectrode space contains superimposed d-c and a-c components which are represented by $\bar{E}_0$ and $\bar{E}_1 \sin \omega t$, respectively, where $\bar{E}_0$ and $\bar{E}_1$ are functions of space coordinates only. The resultant electric intensity is

$$\bar{E} = \bar{E}_0 + \bar{E}_1 \sin \omega t \qquad (2)$$

By inserting the potential from Eq. (1) and the electric intensity from (2) into (3.04-8), we obtain an expression for the instantaneous power transfer,

$$p = -e\bar{v} \cdot (\bar{E}_0 + \bar{E}_1 \sin \omega t) = i_c(V_0 + V_1 \sin \omega t) \qquad (3)$$
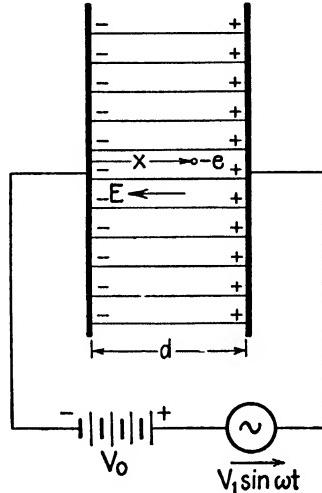
Since the d-c and a-c fields both exert a force upon the electron, there will be power transfer from both potential sources to the electron. Let p_{dc} and p_{ac} represent the instantaneous power transfer to the electron from the sources of d-c and a-c potential, respectively. We may separate the d-c and a-c power terms in Eq. (3) and write

$$p_{dc} = -e\bar{v}\cdot\bar{E}_0 = i_c V_0 \tag{4}$$

$$p_{ac} = -e\bar{v}\cdot\bar{E}_1 \sin \omega t = i_c V_1 \sin \omega t \tag{5}$$

$$p = p_{dc} + p_{ac} \tag{6}$$

The induced current may be obtained either from Eq. (4) or (5), yielding

$$i_c = -\frac{e\bar{v}\cdot\bar{E}_0}{V_0} = -\frac{e\bar{v}\cdot\bar{E}_1}{V_1} \tag{7}$$

It should be noted that the electron velocity $\bar{v}$ is dependent upon both the d-c and a-c fields.

In Eqs. (4) and (5) the terms containing the velocity and electric intensity express the phenomena occurring within the tube, whereas the terms containing voltage and current express the relationships in the external circuit.

In most vacuum-tube applications, a majority of the electrons move in such a manner as to be accelerated by the d-c field but retarded by the a-c field. Hence, these electrons take energy from the source of d-c potential and transfer a portion of this energy to the source of a-c potential. The a-c potential usually results from an alternating current (induced current plus capacitive current) flowing through the load impedance in the external circuit, which produces an a-c potential drop across this impedance. Hence, we shall refer to the load impedance as the source of a-c potential. The electrons, therefore, serve as an intermediary, taking energy from the source of d-c potential and transferring a portion of this, in the form of a-c energy, to the load impedance.

It is of course necessary to control the flow of electrons so that a majority of them flow through the alternating field during its retarding phase. The grid in the ordinary triode tube serves this purpose. The d-c and a-c components of an electric field may be either superimposed as in the triode and magnetron tubes, or they may exist in separate parts of the tube, as in the klystron.

The gain in kinetic energy of a single electron may be obtained by inserting $\bar{E} = \bar{E}_0 + \bar{E}_1 \sin \omega t$ in Eq. (3.04-10), thus

$$\tfrac{1}{2}m(v_b^2 - v_a^2) = -e\int_a^b \bar{E}_0\cdot d\bar{l} - e\int_a^b (\bar{E}_1 \sin \omega t)\cdot d\bar{l} \tag{8}$$

If points a and b are at the cathode and plate, respectively, then the first integral in Eq. (8) is the d-c potential difference and this equation becomes

$$\tfrac{1}{2}m(v_b^2 - v_a^2) = eV_0 - e\int_a^b (\bar{E}_1 \sin \omega t)\cdot d\bar{l} \qquad (9)$$

The two terms on the right-hand side of Eq. (9) represent the contributions to the gain in kinetic energy of the electron due to the d-c and a-c fields. A positive sign in either term (after it has been evaluated) signifies energy transfer from the corresponding potential source to the electron, whereas a negative sign denotes energy transfer in the reverse direction.

The conversion efficiency η_e for the complete travel of a *single electron* is the ratio of a-c energy output to d-c energy input, or

$$\eta_e = \frac{1}{V_0}\int_a^b (\bar{E}_1 \sin \omega t)\cdot d\bar{l} \qquad (10)$$

The sign of the a-c energy has been reversed in Eq. (10) in order that power output and power input both have positive signs, yielding a positive efficiency.

If the electron transit angle is small, the field is practically stationary during the flight of any one electron. For the purpose of evaluating the integral term of Eq. (9), we may treat $\sin \omega t$ as a constant and replace the integral term by $-e\int_a^b (\bar{E}_1 \sin \omega t)\cdot d\bar{l} = eV_1 \sin \omega t$, giving

$$\tfrac{1}{2}m(v_b^2 - v_a^2) = e(V_0 + V_1 \sin \omega t) \qquad (11)$$

with a corresponding efficiency

$$\eta_e = -\frac{V_1 \sin \omega t}{V_0} \qquad (12)$$

Maximum power output and maximum efficiency occur when $\sin \omega t = -1$, *i.e.*, when the a-c potential has its maximum retarding value.

In order to obtain 100 per cent conversion efficiency, the retarding a-c field must have sufficient magnitude to just bring the electron to rest at the end of its journey (point b). All of the energy taken from the d-c source is then transferred to the a-c potential source (the load impedance), and the terminal energy of the electron, as given by Eq. (9), is zero. However, this may present an anomalous situation since, if the superimposed d-c and a-c fields exert equal and opposite forces on an electron (which is assumed to have zero initial velocity), then the electron will never start in motion. Consequently there is no power output, although the theoretical efficiency would be 100 per cent. In general, increasing the retard-

ing a-c field has the effect of reducing the electron velocity and increasing the electron transit time.

The electron transit time (through the a-c field) can be greatly reduced if the d-c and a-c fields are in separate parts of the tube (as in the klystron). Electrons are then accelerated by the d-c field before entering the a-c field. The electrons enter the retarding a-c field with a high velocity. This results in a relatively large power output per electron and small transit time.

When a charge moves through an electrostatic field, the sum of the kinetic and potential energies of the charge remains constant. However, this principle is not always true in time-varying fields. If the field varies with respect to time during the electron's flight, then the sum of the kinetic and potential energies of an electron at any point in its flight is equal to the sum of the kinetic and potential energies at the beginning of the flight *plus an additional term representing the change in energy of the charge due to the time variation of the field.*

3.06. Power Transfer Resulting from Space-charge Flow.—The concepts of induced current and power transfer may now be extended so as to include the effects of a large number of charges in motion. Consider a diode of arbitrary geometry having an electric intensity $\bar{E}$ and a space-charge density q_τ. Both $\bar{E}$ and q_τ are functions of space coordinates and time. It is assumed that the space-charge density is small enough that the field which it produces is negligible in comparison with the field resulting from the potentials on the tube electrodes.[1]

The force experienced by a charge $q_\tau \, d\tau$ occupying a differential volume $d\tau$ is

$$d\bar{f} = \bar{E} q_\tau \, d\tau \tag{1}$$

Inserting this force into Eq. (3.04-6), we obtain the power transfer from the field to the differential space-charge element as $dp = q_\tau \bar{E} \cdot (d\bar{l}/dt) \, d\tau$. Integration over the entire volume τ of the diode yields the instantaneous power transfer from the field to the space charge,

$$p = \int_\tau q_\tau \bar{v} \cdot \bar{E} \, d\tau = \int_\tau \bar{J}_c \cdot \bar{E} \, d\tau \tag{2}$$

where $\bar{J}_c = q_\tau \bar{v}$ is the convection current density. The quantity $\bar{J}_c \cdot \bar{E}$ in the last term of Eq. (2) represents the power transfer per unit volume. If the space charge consists of electrons, q_τ is negative.

The moving space charge induces a current i_c in the external circuit. If the potential difference between electrodes is V, the power transfer from the source of potential to the field is $p = i_c V$. Equating this to the power

---

[1] The effect of the field set up by the space charge is to cause mutual power transfer between the charges themselves.

received by the space charge, we obtain

$$p = \int_\tau \mathbf{J}_c \cdot \bar{E}\, d\tau = i_c V \tag{3}$$

$$i_c = \frac{1}{V} \int_\tau \mathbf{J}_c \cdot \bar{E}\, d\tau \tag{4}$$

For the parallel-plane diode, we have $E = V/d$ and $d\tau = A\, dx$. Equations (3) and (4), in scalar form, then become

$$p = \frac{VA}{d} \int_0^d J_c\, dx = i_c V \tag{5}$$

$$i_c = \frac{A}{d} \int_0^d J_c\, dx \tag{6}$$

Now assume that the potential difference is $V = V_0 + V_1 \sin \omega t$ and let the electric intensity be represented by $\bar{E} = \bar{E}_0 + \bar{E}_1 \sin \omega t$ where $\bar{E}_0$ and $\bar{E}_1$ are again functions of space coordinates. Inserting these into Eq. (3) and separating the d-c and a-c power terms, as in Eqs. (3.05-4 and 5), we obtain

$$p_{dc} = \int_\tau \mathbf{J}_c \cdot \bar{E}_0\, d\tau = i_c V_0 \tag{7}$$

$$p_{ac} = \int_\tau \mathbf{J}_c \cdot \bar{E}_1 \sin \omega t\, d\tau = i_c V_1 \sin \omega t \tag{8}$$

$$p = p_{dc} + p_{ac} \tag{9}$$

The time-average d-c and a-c power may be obtained by integrating Eqs. (7) and (8) over a complete cycle and dividing by 2π. Applying this procedure to the last terms in Eqs. (7) and (8), we obtain

$$P_{dc} = \frac{1}{2\pi} \int_0^{2\pi} i_c V_0\, d(\omega t) \tag{10}$$

$$P_{ac} = \frac{1}{2\pi} \int_0^{2\pi} i_c V_1 \sin \omega t\, d(\omega t) \tag{11}$$

The efficiency is then

$$\eta = \left| \frac{P_{ac}}{P_{dc}} \right| = \frac{\displaystyle\int_0^{2\pi} i_c V_1 \sin \omega t\, d(\omega t)}{\displaystyle\int_0^{2\pi} i_c V_0\, d(\omega t)} \tag{12}$$

In general, the current i_c contains a d-c component, a fundamental a-c component, and higher harmonic a-c components. Only the d-c component of current contributes to the time-average d-c power, and only the fundamental a-c component contributes to the a-c power output (since the a-c voltage was assumed to have no higher harmonics). If we represent the induced current by

$$i_c = I_0 + I_1 \sin (\omega t + \alpha_1) + I_2 \sin (2\omega t + \alpha_2) + \cdots \qquad (13)$$

then Eqs. (10), (11), and (12) become

$$P_{dc} = I_0 V_0 \qquad (14)$$

$$P_{ac} = \tfrac{1}{2} I_1 V_1 \cos \alpha_1 \qquad (15)$$

The conversion efficiency for the tube alone is

$$\eta = \frac{\tfrac{1}{2} I_1 V_1 \cos \alpha_1}{I_0 V_0} \qquad (16)$$

It should be noted that V_0 and V_1 are the potentials *at the tube electrodes*. The d-c supply voltage V_b is equal to the potential V_0 at the tube terminals plus the d-c voltage drop in the load resistance, *i.e.*, $V_b = V_0 + I_0 R_L$. Hence the total power supplied by the d-c source is $P_{dc} = V_b I_0$. The corresponding conversion efficiency (including the loss in R_L) is $\eta = \tfrac{1}{2} V_1 I_1 \cos \alpha_1 / V_b I_0$.

3.07. Example of Power and Energy Transfer.—As an illustration of the foregoing concepts, consider the power and energy transfer in the class C oscillator of Fig. 6.

The potential difference between the cathode and plate contains a d-c component and an a-c component. The a-c potential is developed across the load impedance and is due to the a-c component of current flowing through this impedance. The grid potential serves as an electron gate, allowing electrons to flow into the grid-plate region during the retarding phase of the alternating field only. Thus, in Fig. 6b, the plate current (induced current in the external circuit) flows only when the alternating potential is negative.

The principal energy transfer occurs in the grid-plate region of the tube, since this is the region of high electric intensity and high electron velocity. As the electrons travel from grid to plate, they are accelerated by the d-c field and retarded by the a-c field; hence they take energy from the source of d-c potential and transfer a portion of it, in the form of a-c energy, to the load impedance. The difference between the energy gained by an electron from the d-c source and that transferred to the load impedance is its residual kinetic energy. This is dissipated when the electron strikes the plate. For maximum efficiency, the conduction angle of the current should

be as small as possible and the retarding a-c field should be as large as possible, provided that it is not large enough to introduce appreciable transit-time delay.   This assures maximum retardation of the electrons by the a-c field.
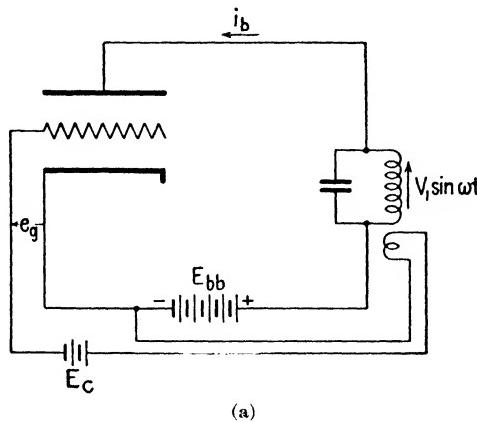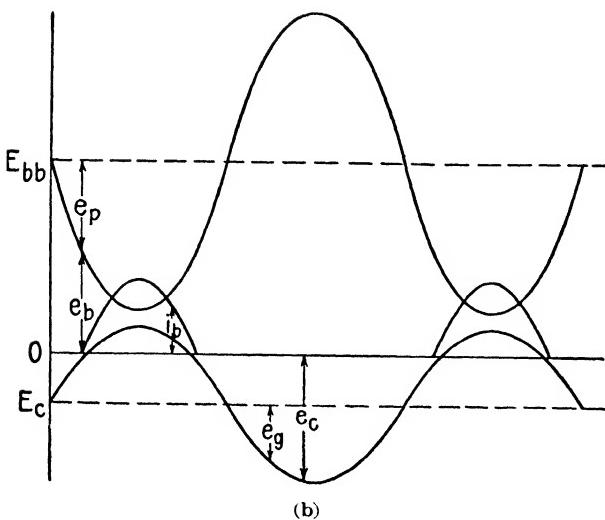
FIG. 6.—Class C oscillator and characteristic curves.

It is entirely possible for electrons to transfer their energy to an external circuit without themselves ever entering this circuit.   Thus, referring to Fig. 7, assume the purely hypothetical case of a space charge oscillating back and forth between two parallel planes at a frequency equal to the antiresonant frequency of the tuned circuit.   The motion of the electrons induces an alternating current in the *L-C* circuit which, in turn, produces

an a-c potential difference between the diode planes.   If the electrons oscillate in such a phase that they are always retarded by the a-c field, they will transfer part of their energy to the *L-C* circuit during each cycle of oscillation even though the electrons may never enter this circuit.   It
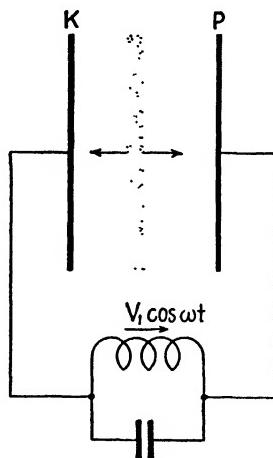
Fig. 7.—Energy transfer resulting from the motion of an oscillating space charge.

would of course be necessary to find a means of supplying energy to the space charge if it is to continue to oscillate.   The mechanism of oscillation described here is similar to that of the positive-grid oscillator.

PROBLEMS

1. Using the equation of continuity, show that the convection-current density is independent of distance in a parallel-plane diode having a d-c applied potential.   Show that the space-charge density varies inversely with electron velocity.

2. A parallel-plane diode is operated with temperature-limited emission.   The diode planes each have an area of 5 sq cm and are 0.5 cm apart.   A d-c potential difference of 500 volts produces a current of 200 ma.
   (a) Obtain an expression for the electron velocity as a function of distance assuming zero emission velocity.
   (b) Compute the total charge enclosed between the diode planes, the convection-current density, and the charge density (see Prob. 1).
   (c) Compute the data and plot curves of q_T, v, J_c, and power transfer per unit volume ($J_c E$) as functions of distance.

CHAPTER 4

THE PHYSICAL BASIS OF EQUIVALENT CIRCUITS

The analysis of vacuum-tube circuits is greatly simplified if an equivalent circuit can be found to represent the tube. This equivalent circuit must yield the same relationship between the applied voltages and currents at its terminals as the tube which it represents.

There are two important methods of deriving such equivalent circuits. The simpler and more common method might be called the Taylor's series or empirical method. It gives the conventional equivalent circuit which is ordinarily used in the analysis of vacuum-tube circuits. In this method, statically determined tube characteristics are represented in the region about a reference or operating point by a Taylor's series. Then the variation in plate current from the operating point value is assumed to be representable by the linear term of the series. In this way the plate-current variation is related to the variations in electrode voltages from their operating point values. Constant-voltage or constant-current equivalent generators can be used to represent the plate-current variations. Interelectrode capacitances are treated as parts of the circuits outside the tube. In this method little attention is given to the electronic phenomena occurring within the tube. Such equivalent circuits are satisfactory for solving problems for which the assumption of linearity between electrode voltage and current variations is satisfactory and for which the frequencies of the applied voltages are such that the transit time of the electrons is negligible in comparison with the period of the voltage wave.

In the second method, however, the expressions for voltage and current are derived from basic dynamical relationships which represent the behavior of the electrons within the tube. These relationships are then used to obtain the equivalent circuits of diodes. The equivalent circuits for triodes and other multielectrode tubes are synthesized from the equivalent circuits for diodes. Since electron transit-time effects are taken into consideration, these equivalent circuits may be used at microwave frequencies. In general, the circuit elements in these equivalent circuits are functions of electron transit time. This method of deriving equivalent circuits will be called the electronic method.

In this chapter we will derive the conventional equivalent circuit of the triode tube by the Taylor's series method. Then, using the electronic

method, the equivalent circuit for the temperature-limited diode will be derived and the derivation of the equivalent circuit for the space-charge-limited diode will be outlined. The method of synthesizing the equivalent circuits for triode and tetrode tubes from diodes will be indicated.

4.01. Conventional Equivalent Circuit of the Triode Tube.—A triode vacuum tube is one with three electrodes. However, tetrodes, pentodes, and other multielectrode tubes may function essentially as triodes if only two of their electrodes in addition to their cathodes have varying poten-
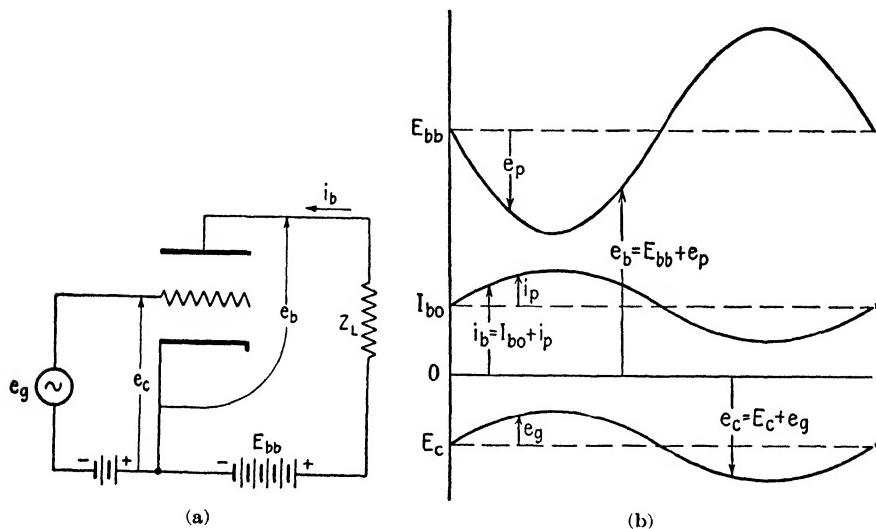
Fig. 1.—Triode amplifier and characteristics for class A operation.

tials. For example, a pentode tube functions as a triode with special characteristics if its screen and suppressor grids are held at a constant potential with respect to its cathode.

Let us briefly consider the conventional equivalent circuit for vacuum tubes operating as triodes at frequencies for which transit-time effects are negligible. In Fig. 1, let i_b, e_c, and e_b represent the instantaneous values of total plate current, grid voltage, and plate voltage, respectively.[1] The varying components of grid and plate voltages and plate current are e_g, e_p, and i_p, respectively. In Fig. 1b these varying components are shown for illustrative purposes with sinusoidal waveforms and with proper phase relationships for a resistive load impedance.

The plate current consists of the quiescent or reference-point value I_{b0} plus the variational component i_p, thus $i_b = I_{b0} + i_p$. The instantaneous

---

[1] The bold-faced **E** in this chapter and in Chap. 5 is used to denote vacuum tube voltages, with subscripts in conformity with I.R.E. standards.

value of plate current may be expressed in terms of a Taylor's series as follows:

$$i_b = I_{b0} + \left(e_g \frac{\partial}{\partial e_c} + e_p \frac{\partial}{\partial e_b}\right) i_b + \frac{1}{2!}\left(e_g \frac{\partial}{\partial e_c} + e_p \frac{\partial}{\partial e_b}\right)^2 i_b +$$

$$\cdots \frac{1}{n!}\left(e_g \frac{\partial}{\partial e_c} + e_p \frac{\partial}{\partial e_b}\right)^n i_b + \cdots \quad (1)$$

In this representation of the series, e_p is the variational component of plate voltage, e_g is the variational component of grid voltage, and the partial-derivative operators operate only on i_b or its derivatives.
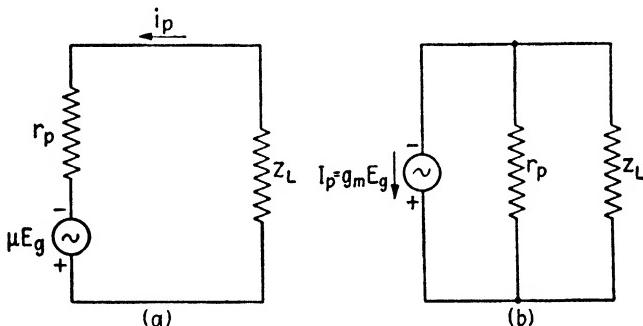
FIG. 2.—Equivalent plate circuits of a triode.

We now define a transconductance g_m, a plate resistance r_p, and an amplification factor μ as follows:

$$g_m = \frac{\partial i_b}{\partial e_c} \qquad r_p = \frac{1}{\partial i_b / \partial e_b} \qquad \mu = g_m r_p \quad (2)$$

If μ can be considered constant, we can write Eq. (1) in the form

$$i_p = g_m\left(e_g + \frac{e_p}{\mu}\right) + \frac{1}{2!}\frac{\partial g_m}{\partial e_c}\left(e_g + \frac{e_p}{\mu}\right)^2 + \cdots \frac{1}{n!}\frac{\partial^{n-1} g_m}{\partial e_c^{n-1}}\left(e_g + \frac{e_p}{\mu}\right)^n + \cdots \quad (3)$$

Hence, i_p can be represented as a power series in terms of the voltage $[e_g + (e_p/\mu)]$. The second- and higher-order terms yield distortion components of i_p. By retaining only the linear term in Eq. (3) and substituting (for a resistive load)

$$e_p = -i_p R_L \quad (4)$$

we obtain

$$i_p = \frac{\mu e_g}{r_p + R_L} \quad (5)$$

For sinusoidally varying quantities, Eq. (5) can be expressed in terms of the complex voltage E_g, current I_p, and load impedance Z_L, yielding

$$I_p = \frac{\mu E_g}{r_p + Z_L} \tag{6}$$

Equation (6) may be represented by either of the equivalent circuits shown in Fig. 2. The circuit of Fig. 2a contains a series combination of a constant-voltage generator whose voltage is μE_g, a plate resistance r_p, and a load impedance Z_L. Figure 2b contains the plate resistance and load impedance in parallel with a constant-current generator whose current is $g_m E_g$.
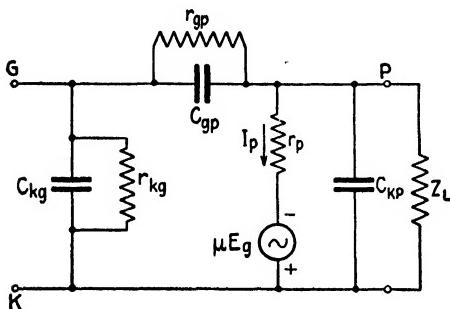
Fig. 3.—Equivalent circuit of a negative-grid triode, including interelectrode capacitances and losses.

If the interelectrode capacitances and the losses in the cathode-grid and grid-plate circuit are included, the equivalent circuit takes the form shown in Fig. 3. At high frequencies the lead inductances are important.[1,2] These may be added in series with the cathode. grid, and plate leads to complete the equivalent circuits for moderately high frequencies.

4.02. Procedure in Solving the Temperature-limited Diode.—The characteristics of any vacuum tube are dependent upon the total effect of the electrons moving in the interelectrode space. In the electronic method, the voltages and currents at the tube electrodes are expressed in terms of the equations for the motion of electrons within the tube. These expressions may be used to determine the current flowing for given applied voltages, or conversely, to determine the voltage required to produce a given current. The relationships between the a-c components of voltage and current determine the values of the impedance or admittance elements in the equivalent circuits.

Consider the case of the parallel-plane diode of Fig. 4, which is assumed to have temperature-limited current. In general, the induced plate current i_b in the external circuit resulting from the motion of electrons in the interelectrode space of the diode may be represented by a Fourier series of the form

$$i_b = \iota I_{b0} + I_1 \sin (\omega t + \alpha_1) + I_2 \sin (2\omega t + \alpha_2) + \cdots I_n \sin (n\omega t + \alpha_n) \tag{1}$$

[1] LLEWELLYN, F. B., Equivalent Networks of Negative-grid Vacuum Tubes at Ultra High Frequencies, *Bell System Tech. J.*, vol. 15, pp. 575–86; October, 1936.

[2] STRUTT, M. J. O., and A. VAN DER ZIEL, The Causes for the Increase of the Admittances of Modern High-frequency Amplifier Tubes on Short Waves, *Proc. I.R.E.*, vol. 26, pp. 1011–1032; August, 1938.

A question might arise as to how it is possible to have a-c components of plate current if this current is temperature limited. Under temperature-limited conditions, all of the emitted electrons are drawn to the plate and low-frequency variations in applied voltage do not alter the rate at which electrons arrive at the plate. However, at microwave frequencies, the electron transit angle may be quite large and the electrons therefore remain in the interelectrode space for an appreciable portion of the a-c cycle. Electrons which are emitted from the cathode during the retarding phase of the a-c voltage are slowed down, whereas those emitted during the accelerating phase of a-c voltage are speeded up. The faster electrons tend to overtake the slower electrons which were emitted at an earlier phase and there is consequently a tendency toward bunching of the electrons in the interelectrode space. This bunching effect produces the alternating components of plate current.

FIG. 4.—Parallel-plane diode with d-c and a-c applied potentials.

Throughout the following analysis, small-amplitude operation will be assumed, i.e., $V_1 \ll V_0$. The harmonic terms in Eq. (1) are then small in comparison with the fundamental component. If we discard the harmonic components of current, leaving the d-c and fundamental a-c components, we have

$$i_b = I_{b0} + I_1 \sin (\omega t + \alpha_1)$$

which can be written

$$i_b = I_{b0} + I_1' \sin \omega t + I_1'' \cos \omega t \tag{2}$$

where I_1' is the fundamental component of current in phase with the a-c voltage and I_1'' is the quadrature component.

The d-c conductance g_0, a-c conductance g_1, and a-c susceptance b_1, are then

$$g_0 = \frac{I_{b0}}{V_0} \qquad g_1 = \frac{I_1'}{V_1} \qquad b_1 = \frac{I_1''}{V_1} \tag{3}$$

The values of the a-c conductance g_1 and susceptance b_1 are dependent upon the electron transit angle and they may be either positive or negative.

In complex form, the a-c admittance is

$$Y = g_1 + jb_1 \tag{4}$$

This is the expression for the admittance of a parallel combination of conductance g_1 and susceptance b_1. Since this admittance represents only the electronic effects, another parallel branch representing the susceptance, b_c, due to the interelectrode capacitance, must be added to obtain the complete equivalent circuit. The resulting equivalent circuit is shown in

Fig. 5. The susceptance b_1 may be either capacitive or inductive depending upon the electron transit angle.
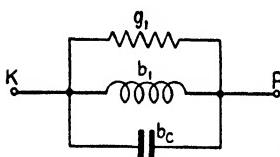
FIG. 5.—A-c equivalent circuit of a temperature-limited diode. Values of g_1 and b_1 are given in Fig. 6.

In order to evaluate the admittances in Eq. (3), it is necessary to obtain expressions for I_0, I_1', and I_1'' in terms of the applied potentials. The general method of analysis presented here was developed by Benham, Llewellyn and North.[1-7]

4.03. Equivalent Circuit of the Temperature-limited Diode.—In the parallel-plane diode with temperature-limited current, let N be the number of electrons emitted from the cathode per second. The charge emitted between time t_0 and $t_0 + dt_0$ is $Ne\,dt_0$. By substituting this charge in Eq. (3.04-12) in place of e, we obtain the induced current $di_b = Nev\,dt_0/d$. The total induced current at any instant of time t is due to electrons emitted from the cathode between time $t_0 = t - T_1$ (since T_1 is the total transit time these electrons are arriving at the plate at time t) and $t_0 = t$ (these electrons are leaving the cathode at time t). The total induced current is, therefore,

$$i_b = \int_{t_0=t-T_1}^{t_0=t} \frac{Nev}{d}\,dt_0 \tag{1}$$

The electron velocity, given by Eq. (2.07-4), with $v_0 = 0$, is now substituted into Eq. (1). The current is being evaluated at a given instant of

[1] BENHAM, W. E., Theory of the Internal Action of Thermionic Systems at Moderately High Frequencies, *Phil. Mag.*, part I, vol. 5, pp. 641–662; March, 1928; part II, vol. 7, pp. 457–515; February, 1931.

[2] BENHAM, W. E., A Contribution to Tube and Amplifier Theory, *Proc. I.R.E.*, vol. 26, pp. 1093–1170; September, 1938.

[3] LLEWELLYN, F. B., Vacuum Tube Electronics at Ultra-high Frequencies, *Proc. I.R.E.*, vol. 21, no. 11, pp. 1532–1572; November, 1933.

[4] LLEWELLYN, F. B., Note on Vacuum Tube Electronics at Ultra-high Frequencies, *Proc. I.R.E.*, vol. 23, no. 2, pp. 112–127; February, 1935.

[5] LLEWELLYN, F. B., "Electron Inertia Effects," Cambridge University Press, London, 1941.

[6] LLEWELLYN, F. B., and L. C. PETERSON, Vacuum Tube Networks, *Proc. I.R.E.*, vol. 32, no. 3, pp. 144–166; March, 1944.

[7] NORTH, D. O., Analysis of the Effects of Space Charge on Grid Impedance, *Proc. I.R.E.*, vol. 24, no. 1, pp. 108–136; January, 1936.

time t; hence t is held constant in the integration of Eq. (1) and t_0 is the variable. Integrating and forming the ratio i_b/Ne, we obtain

$$
\frac{i_b}{Ne} = \frac{e}{md^2} \int_{t_0=t-T_1}^{t_0=t} \left[V_0(t - t_0) - \frac{V_1}{\omega} (\cos \omega t - \cos \omega t_0) \right] dt_0
$$

$$
= \frac{e}{md^2} \left\{ \frac{V_0 T_1^2}{2} + V_1 T_1^2 \left[\sin \omega t \left(\frac{1 - \cos \theta}{\theta^2} \right) + \cos \omega t \left(\frac{\sin \theta - \theta}{\theta^2} \right) \right] \right\} \quad (2)
$$

where θ is the electron transit angle and is given by

$$
\theta = \omega T_1 \tag{3}
$$

We wish to separate the d-c and a-c components of current. The term $V_0 T_1^2 / 2$ in Eq. (2) contains both d-c and a-c components, since the total transit time T_1 is a function of t. To separate the d-c and a-c components, it is necessary to return to Eq. (2.07-5) which expresses the displacement of the electron. Substituting $x = d$, $T = T_1$, $t_0 = t - T_1$, and $\theta = \omega T_1$, with $v_0 = 0$, in Eq. (2.07-5), we obtain

$$
1 = \frac{e}{md^2} \left\{ \frac{V_0 T_1^2}{2} + V_1 T_1^2 \left[\sin \omega t \left(\frac{\cos \theta - 1 + \theta \sin \theta}{\theta^2} \right) \right.\right.
$$
$$
\left.\left. - \cos \omega t \left(\frac{\sin \theta - \theta \cos \theta}{\theta^2} \right) \right] \right\} \quad (4)
$$

Subtracting Eq. (4) from (2), we complete the separation process,

$$
\frac{i_b}{Ne} = 1 + \frac{eV_1 T_1^2}{md^2} \left\{ \sin \omega t \left[\frac{2(1 - \cos \theta) - \theta \sin \theta}{\theta^2} \right] \right.
$$
$$
\left. + \cos \omega t \left[\frac{2 \sin \theta - \theta(1 + \cos \theta)}{\theta^2} \right] \right\} \quad (5)
$$

Now assume that the total transit time is approximately equal to the d-c transit time, *i.e.*, $T_1 \approx T_0$. Equation (2.07-11) gives the d-c transit time

$$
T_0 = d \sqrt{\frac{2m}{eV_0}} \tag{2.07-11}
$$

If T_0 is substituted for T_1 in Eq. (5), the current ratio becomes

$$
\frac{i_b}{Ne} = 1 + \frac{2V_1}{V_0} \left\{ \sin \omega t \left[\frac{2(1 - \cos \theta) - \theta \sin \theta}{\theta^2} \right] \right.
$$
$$
\left. + \cos \omega t \left[\frac{2 \sin \theta - \theta(1 + \cos \theta)}{\theta^2} \right] \right\} \quad (6)
$$

Comparing this expression with Eq. (4.02-2), we find that

$$I_{b0} = Ne$$

$$I_1' = 2I_{b0} \frac{V_1}{V_0} \left[\frac{2(1 - \cos \theta) - \theta \sin \theta}{\theta^2} \right] \tag{7}$$

$$I_1'' = 2I_{b0} \frac{V_1}{V_0} \left[\frac{2 \sin \theta - \theta(1 + \cos \theta)}{\theta^2} \right]$$
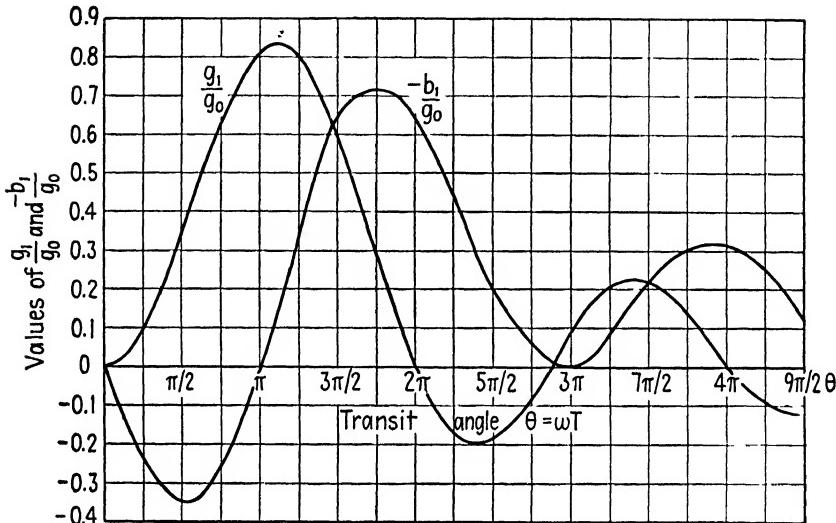
FIG. 6.—A-c conductance and susceptance of the temperature-limited diode, as a function of transit angle. The susceptance is capacitive for small transit angles.

By substituting these results in Eq. (4.02-3), we obtain the final equations for the admittance parameters,

$$g_0 = \frac{Ne}{V_0} = \frac{I_{b0}}{V_0}$$

$$\frac{g_1}{g_0} = 2 \left[\frac{2(1 - \cos \theta) - \theta \sin \theta}{\theta^2} \right] = \frac{\theta^2}{6} \left[1 - \frac{\theta^2}{15} + \cdots \right] \tag{8}$$

$$\frac{b_1}{g_0} = 2 \left[\frac{2 \sin \theta - \theta(1 + \cos \theta)}{\theta^2} \right] = \frac{\theta}{3} \left[1 - \frac{3\theta^2}{20} + \cdots \right]$$

Values of g_1/g_0 and b_1/g_0 are plotted in Fig. 6 as functions of the transit angle θ. As the transit angle increases, the conductance is alternately positive and negative with a maximum near $\theta = \pi$. Negative values of conductance signify energy transfer from the electron stream to the source of

alternating potential.   Hence the temperature-limited diode can be used as an oscillator at frequencies for which its a-c conductance is negative.

The susceptance b_1 is alternately capacitive and inductive as the transit angle increases.   By adding the susceptance b_c, due to the static capacitance between the electrodes, to b_1 we obtain the total susceptance.   Since the static capacitance is $C = \epsilon A/d$ farads, the corresponding susceptance is

$$b_c = \frac{\omega \epsilon A}{d} \tag{9}$$

In this equation, A is the area of one of the plane electrodes.

The series form of the above equations is convenient for evaluating the admittance for small transit angles.   Substituting $\theta = \omega T_1$, and taking the first terms only in each of the above series expansions, we obtain, for small transit angles,

$$\frac{g_1}{g_0} = \frac{(2\pi f)^2 T_0^2}{6}$$

$$\frac{b_1}{g_0} = \frac{2\pi f T_0}{3} \tag{10}$$

Hence, the conductance is proportional to the square of frequency and transit time, whereas the susceptance is proportional to the first power of frequency and transit time.   For small values of transit angle, the conductance term is small in comparison with the susceptance and the admittance is approximately equal to the susceptance b_1.   These are the conclusions obtained by Ferris, by a somewhat different method of analysis.[1]

4.04. Relationships for the Space-charge-limited Diode.—Consider a parallel-plane diode with arbitrary applied potentials and appreciable space charge between the diode planes.   The current density at any point in the interelectrode space is the sum of the convection-current density $q_\tau \bar{v}$ and the displacement-current density $\epsilon(\partial \bar{E}/\partial t)$, as given by Eq. (3.02-5), thus [2]

$$J = q_\tau \bar{v} + \epsilon \frac{\partial \bar{E}}{\partial t} \tag{1}$$

For the parallel-plane diode, the divergence equation (2.05-7) becomes

$$\frac{\partial E_x}{\partial x} = \frac{q_\tau}{\epsilon} \tag{2}$$

[1] FERRIS, W. R., Input Resistance of Vacuum Tubes as Ultra-high-frequency Amplifiers, *Proc. I.R.E.*, vol. 24, pp. 82–107; January, 1936.
[2] The displacement current density includes both the capacitive current and the displacement current due to electron motion in the diode.

By substituting q_r from this equation and $v = dx/dt$ into Eq. (1), written in one-dimensional form, we obtain

$$J = \epsilon \left(\frac{\partial E_x}{\partial x} \frac{dx}{dt} + \frac{\partial E_x}{\partial t} \right) \qquad (3)$$

The electric intensity is a function of x and t. According to the rules of the partial differentiation of a function of two variables, Eq. (3) is equivalent to

$$J = \epsilon \frac{dE_x}{dt} \qquad (4)$$

Here we have an interesting relationship which states that the current density is proportional to the time rate of increase of electric intensity *as experienced by the moving electron*. In other words, from the point of view of the moving electron, there can be only a displacement current. This is in agreement with the discussion at the end of Sec. 3.03.

The acceleration of an electron in an electric field is given by the first of Eq. (2.06-3)

$$\frac{d^2x}{dt^2} = - \frac{eE_x}{m} \qquad (2.06\text{-}3)$$

Substitution of the electric intensity from this equation into Eq. (4) yields

$$\frac{d^3x}{dt^2} = - \frac{eJ}{\epsilon m} \qquad (5)$$

This equation states that the current density is proportional to the time rate of change of electron acceleration.

The potential may likewise be related to the equations of electron motion. Starting with $V = - \int_0^d E \, dx$ and substituting E from Eq. (2.06-3), we obtain

$$V = \frac{m}{e} \int_0^d \frac{d^2x}{dt^2} \, dx \qquad (6)$$

The foregoing relationships are valid for a parallel-plane diode having any degree of space charge.

4.05. Space-charge-limited Diode with D-C Potential.—Now consider the special case of a parallel-plane diode having a d-c potential difference, and space-charge-limited current. Let $J = J_0$ be the d-c current density. Since the electrons flow in the positive x direction, J_0 will be negative. Substituting $J = J_0$ into Eq. (4.04-5) and performing successive integrations, we obtain the electron acceleration, velocity, and displacement equa-

tions.   Assume that a particular electron leaves the cathode ($x = 0$) at time $t = t_0$, with zero initial velocity.   For complete space-charge-limited current, the electric intensity at the cathode is zero.   Integration of Eq. (4.04-5), with the above substitutions, then yields

$$\frac{d^2x}{dt^2} = -\frac{eJ_0}{\epsilon m}(t - t_0) \tag{1}$$

$$\frac{dx}{dt} = -\frac{eJ_0}{2\epsilon m}(t - t_0)^2 \tag{2}$$

$$x = -\frac{eJ_0}{6\epsilon m}(t - t_0)^3 \tag{3}$$

Let $T = t - t_0$ be the time required for the electron to travel a distance x.   The total electron transit time (cathode to plate) is found by setting $x = d$ and $t - t_0 = T_0$ in Eq. (3), yielding

$$T_0 = \left(-\frac{6\epsilon m d}{eJ_0}\right)^{\frac{1}{3}} \tag{4}$$

The potential required to produce the assumed current density may be obtained by inserting d^2x/dt^2 from Eq. (1) and dx from Eq. (2) into (4.04-6).   However, we first write Eq. (2) in the form $dx/dt = dx/dT = -eJ_0T^2/2\epsilon m$ and then substitute $dx = (-eJ_0T^2/2\epsilon m)\,dT$ into Eq. (4.04-6).   Since potential is being evaluated at a particular instant of time, t is constant and t_0 (or T) is the variable.   With the above substitutions, Eq. (4.04-6) gives

$$V_0 = \frac{eJ_0^2}{2\epsilon^2 m}\int_0^{T_0} T^3\,dT = \frac{eJ_0^2 T_0^4}{8\epsilon^2 m} \tag{5}$$

Several interesting aspects of the space-charge-limited diode are revealed in Eqs. (4) and (5).   Eliminating J_0 from these two equations, we can obtain the following expressions for the total transit time and transit angle:

$$T_0 = 3d\sqrt{\frac{m}{2eV_0}} = 5.05 \times 10^{-6}\frac{d}{\sqrt{V_0}} \qquad \text{seconds}$$

$$\theta = \omega T_0 = 31.7 \times 10^{-6}\frac{fd}{\sqrt{V_0}} \qquad \text{radians} \tag{6}$$

In these equations d is in meters and V_0 in volts.   Comparison of Eqs. (6) and (2.07-11) shows that the total electron transit time in the space-charge-limited diode is three-halves of the transit time in the temperature-limited diode.

By substituting T_0 from Eq. (6) into Eq. (5) and solving for J_0, we obtain the familiar three-halves power Child's law for space-charge-limited current,

$$J_0 = \frac{4\epsilon}{9d^2} \sqrt{\frac{2e}{m}} V_0^{3/2} = 2.33 \times 10^{-6} \frac{V_0^{3/2}}{d^2} \qquad (7)$$

The d-c resistance of the diode *per unit area* is given by

$$r_0 = \frac{V_0}{J_0} = 0.429 \times 10^6 \frac{d^2}{\sqrt{V_0}} \qquad (8)$$

4.06. Space-charge-limited Diode with D-C and A-C Applied Potentials.—The procedure in solving the space-charge-limited diode with combined d-c and a-c applied potentials is similar to that of Sec. 4.05 except that the assumed current density is of the form

$$J = J_0 + J_1 \sin \omega t \qquad (1)$$

The solution is quite lengthy; therefore we shall merely state the end results. The potential required to produce the assumed current is

$$V = V_0 + 8J_1 r_0 \left\{ \sin \omega t \left[\frac{2(1 - \cos \theta) - \theta \sin \theta}{\theta^4} \right] \right.$$
$$\left. - \cos \omega t \left[\frac{1}{6\theta} + \frac{\theta(1 + \cos \theta) - 2 \sin \theta}{\theta^4} \right] \right\} \qquad (2)$$

where $r_0 = V_0/J_0$ is the d-c resistance of unit area of the diode, and $\theta = \omega T_0$ is the total transit angle.

Equation (2) contains voltage components in phase with the current (the $\sin \omega t$ terms) and quadrature components (the $\cos \omega t$ terms). The equivalent a-c circuit is therefore a series circuit containing resistance and reactance. Dividing the a-c terms of Eq. (2) by J_1, we obtain the resistance and reactance, per unit area,

$$r_1 = 8r_0 \left[\frac{2(1 - \cos \theta) - \theta \sin \theta}{\theta^4} \right] = \frac{2r_0}{3} \left[1 - \frac{\theta^2}{15} + \cdots \right] \qquad (3)$$

$$x_1 = -\frac{4r_0}{3\theta} - 8r_0 \left[\frac{\theta(1 + \cos \theta) - 2 \sin \theta}{\theta^4} \right] = -\frac{2}{3} r_0 \left[\frac{3\theta}{10} - \frac{\theta^3}{84} + \cdots \right] \qquad (4)$$

A positive sign in the reactance term signifies inductive reactance, while a negative sign signifies capacitive reactance. The term $-4r_0/3\theta$ in Eq. (4) is the capacitive reactance between the diode planes without space charge. To show this, write $r_0 = V_0/J_0$. Substitute Eqs. (4.05-4 and 5)

and rearrange to obtain $r_0 = -3T_0 d/4\epsilon$. Multiplying this by $4/3\theta$, we obtain

$$\frac{4r_0}{3\theta} = \frac{d}{\omega\epsilon} = \frac{1}{\omega C} \tag{5}$$

where $C = \epsilon/d$ is the capacitance of unit area of diode planes. Letting $x_c = 1/\omega C$ repre-sent the reactance due to the diode capaci-tance and x_e represent the reactance due to

FIG. 7.—Equivalent circuit of the space-charge-limited diode.

space-charge and transit-time effects per unit area, we obtain for Eq. (4)

$$x_1 = -x_c + x_e \tag{6}$$

The series form of Eq. (3) shows that the low-frequency a-c resistance of the space-charge-limited diode is $r_p = \frac{2}{3}r_0$, *i.e.*, two thirds of the d-c resistance. Figure 7 shows the equivalent circuit for the space-charge-limited diode. Values of r_1/r_p, x_1/r_p, and x_c/r_p are plotted as functions
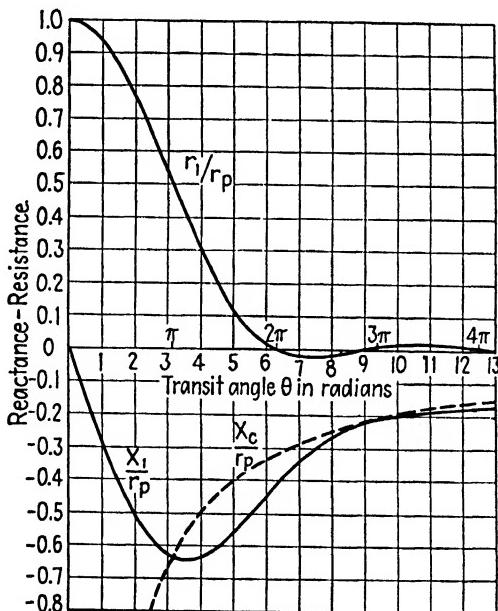
FIG. 8.—Impedance of the space-charge-limited diode as a function of transit angle.

of transit angle $\theta = \omega T_0$ in Fig. 8. The resistance alternates between posi-tive and negative values as the transit angle increases, with a maximum positive value at zero transit angle and maximum negative value at ap-proximately $\theta = 7.2$ radians. Equation (4.05-6) can be used to compute the approximate value of the transit angle, and the corresponding imped-ances for the diode may then be obtained from Fig. 8.

The admittance for the space-charge-limited diode is the reciprocal of its impedance. Figure 9 is a plot of admittance values, showing the low-frequency and high-frequency asymptotes of the susceptance curve. It is interesting to note that the low-frequency capacitance in the equivalent
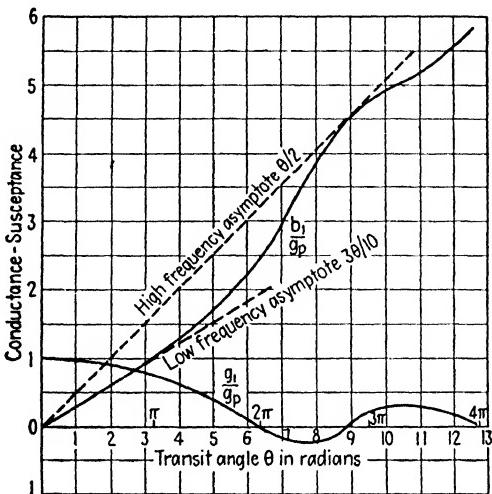
FIG. 9.—Admittance of the space-charge-limited diode as a function of transit angle.

parallel circuit is three fifths of the capacitance without space charge. This effect of space charge in reducing the effective capacitance is sometimes referred to as a reduction in the effective dielectric constant. It is also interesting to observe the similarity between the equations for the impedance of the space-charge-limited diode as given by Eqs. (3) and (4) and the admittance of the temperature-limited diode in Eq. (4.03-8).

FIG. 10.—Equivalent circuit of the triode at microwave frequencies.

4.07. Equivalent Circuit of the Triode.[1]—The triode may be viewed as two adjoining diode regions with a space-charge-limited diode representing the cathode-grid region and a temperature-limited diode for the grid-plate region. The impedance or admittance of the space-charge-limited region may be obtained from Figs. 8 or 9, using the transit angle computed from Eq. (4.05-6). The temperature-limited diode region re-

[1] LLEWELLYN, F. B., and L. C. PETERSON, Interpretation of Ultra-high Frequency Tube Performance in Terms of Equivalent Networks, *Electronic Industries*, vol. 3, pp. 88–90; November, 1944.
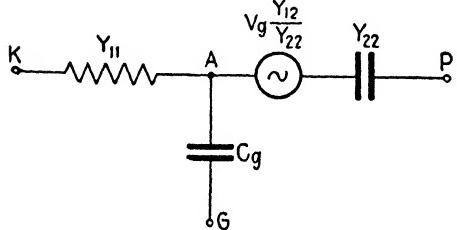
quires a little different treatment from that given in Sec. 4.03, since the initial velocity of the electrons entering the grid-plate region is not zero.

Llewellyn has shown that the equivalent circuit shown in Fig. 10 may be used to represent the triode at all frequencies including microwave frequencies. The admittance elements are determined as follows:

1. Admittance Y_{11} is the admittance of the space-charge-limited diode as obtained from Fig. 9 with the transit angle computed using Eq. (4.05-6).
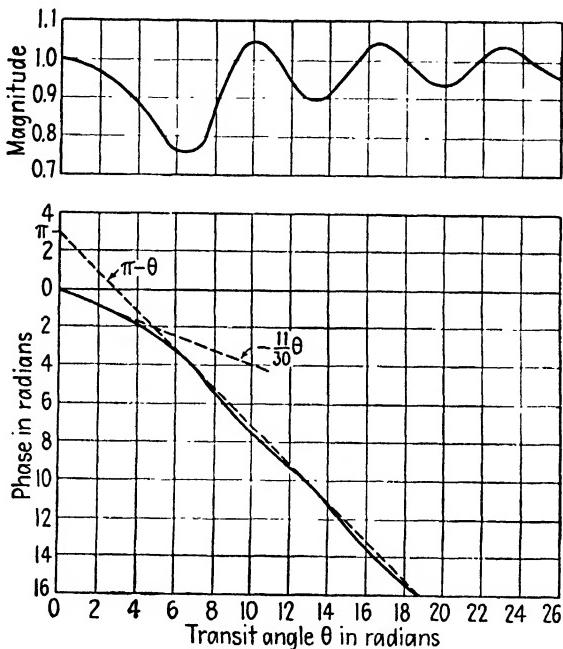
FIG. 11.—Transadmittance of a triode as a function of cathode-grid transit angle.

2. Admittance Y_{22} is approximately equal to the susceptance of the grid-plate capacitance, or $Y_{22} = j\omega C_{gp}$.

3. Point A in Fig. 10 is assumed to be a point midway between grid wires in the grid plane. The capacitance C_g is given approximately by $C_g = \mu_g C_{gp}$, where μ_g is the reciprocal of the screening factor of the tube. Capacitance C_g represents the capacitance between the electron stream at point A and the grid wire.

4. Admittance Y_{12} is given for various values of θ_1 in Fig. 11, where θ_1 is the cathode-grid transit angle. The magnitude of Y_{12} is approximately equal to the conventional transconductance g_m although its phase angle varies with the value of θ_1.

5. The potential difference between the cathode and grid plane is V_1. This is not the same as the potential difference between the cathode and

grid wires. The grid-plane potential may be taken as the potential midway between grid wires. The equivalent generator has a voltage $V_1(Y_{12}/Y_{22})$.

Although the equivalent circuit derived here is valid for all frequencies, numerical difficulties are encountered in attempting to evaluate the admittance elements and equivalent-generator voltage at low frequencies since, as the transit angle decreases, the values of some of the impedances approach the indeterminant form ∞/∞. Consequently, the conventional equivalent circuit of Fig. 3 is preferred where very small transit angles are
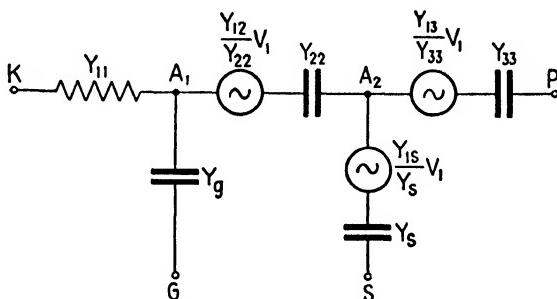
FIG. 12.—Equivalent circuit of a tetrode at microwave frequencies.

involved. The external circuit may be attached to the terminals of the equivalent circuit of Fig. 10, and the complete vacuum-tube circuit solved by the usual methods of circuit analysis.

Equivalent circuits may also be derived for more complicated tube structures using this method. Figure 12 shows the equivalent circuit of a tetrode tube.

PROBLEMS

1. A parallel-plane diode has a space-charge-limited current. The diode planes have an area of 2 sq cm and are separated by a distance of 0.1 cm. The potential difference between cathode and anode is $V = 75 + 10 \sin \omega t$. The d-c current is 10 ma and the frequency is 1,000 megacycles.

   (a) Find the total electron transit time and transit angle (assuming that the a-c potential has negligible effect upon the transit time).

   (b) Determine the d-c plate resistance and the a-c plate resistance.

   (c) Using the transit angle in part (a), evaluate the parameters of the equivalent a-c circuit of the diode.

   (d) At what frequency does the diode resistance have its maximum negative value? What is the value of the resistance at this frequency?

2. A lighthouse tube has parallel-plane electrodes. The cathode-to-grid spacing is 0.014 cm and grid-to-plate spacing is 0.033 cm. The cathode and plate have effective areas of 0.7 sq cm each. The amplification factor and interelectrode capacitances are

$$\mu = 45$$
$$C_{kg} = 2.65 \times 10^{-12} \text{ farad}$$
$$C_{gp} = 1.70 \times 10^{-12} \text{ farad}$$
$$C_{kp} = 0.04 \times 10^{-12} \text{ farad}$$

Assume that the tube is operated with space-charge-limited emission with a potential difference between the cathode and the equivalent grid plane of 3 volts. The potential difference between the cathode and plate is 200 volts. The tube is operating as an oscillator at a frequency of 2,500 megacycles with small-amplitude operation.

(a) Compute the cathode-to-grid transit angle.

(b) Determine the parameters for the equivalent a-c circuit assuming that μ_g is equal to the amplification factor of the tube.

CHAPTER 5

NEGATIVE-GRID TRIODE. OSCILLATORS AND AMPLIFIERS

The conventional triode tube has proven to be an exceedingly versatile tube and is the prototype of many other vacuum tubes in present-day use. By careful design of the tube and the associated circuits, it has been possible to extend the useful range of the triode into the microwave portion of the frequency spectrum. In this range the resonant circuits are usually constructed of transmission-line elements or resonant cavities. The tubes are often designed so that they can be made an integral part of the resonant system.

5.01. Triode Tube Considerations.—Optimum design of the triode tube requires a minimum of

1. Interelectrode capacitance and lead inductance.
2. Electron transit time.

and a maximum of

1. Transconductance.
2. Cathode emission.
3. Plate heat dissipation area.

These are conflicting requirements which necessitate design compromises. Interelectrode capacitance may be reduced by decreasing the physical size of the electrodes and increasing the spacing. However, decreasing the electrode dimensions reduces the maximum safe power dissipation, thereby reducing the power rating of the tube, whereas increasing the interelectrode spacing results in a detrimental increase in electron transit time. Electron transit time may be reduced by the use of a high plate potential. However, the maximum allowable plate potential is determined by the safe plate dissipation of the tube and the breakdown voltage. The lead inductance can be minimized by the use of short thick leads, preferably arranged so that they can become an integral part of the external circuit. Low-loss dielectric seals are used to reduce the dielectric losses.

A useful principle of similitude [1,2] states that if all linear dimensions of a triode are changed by a fixed ratio and the electrode voltages are held constant, the plate current, transconductance, and amplification factor will

[1] LANGMUIR, I., and K. COMPTON, Electrical Discharges in Gases, *Rev. Modern Phys.*, vol. 3, pp. 192–257; April, 1931.
[2] THOMPSON, B. J., and G. M. ROSE, Vacuum Tubes of Small Dimensions for Use at Extremely High Frequencies, *Proc. I.R.E.*, vol. 21, pp. 1707–1721; December, 1933.

remain constant. However, the interelectrode capacitances, electron transit time, and lead inductance vary directly with the linear dimensions. It is apparent, therefore, that some of the undesirable effects encountered at microwave frequencies may be minimized by shrinking all of the linear dimensions of the tube proportionately, although this remedy also reduces the power rating of the tube.

The relatively high input admittance of triode tubes at microwave frequencies imposes a serious limitation upon the operation of the tubes at these frequencies. This is due, in part, to the electron transit-time effects. Consider a triode tube with an alternating potential applied between cathode and grid. If the tube is used as an oscillator or amplifier, most of the electrons flow from cathode to plate during the positive half cycle of grid potential. The electrons are accelerated by the field of the a-c grid potential during their transit through the cathode-grid region and retarded by this field during their transit through the grid-plate region. If transit-time effects are negligible, the electrons take energy from the a-c grid-potential source as they approach the grid, but return an approximately equal amount of energy to this source as they travel from grid to plate. Consequently, there is no net energy transfer from the a-c grid-potential source to the electrons, and the input conductance due to this effect is zero.

However, at microwave frequencies, the electron transit angle may be relatively large. It is then possible for the a-c grid potential to reverse its phase before the electrons reach the plate, so that the electrons are accelerated by the field of the a-c grid potential in both the cathode-grid region and in the grid-plate region. Under these conditions, there is a net energy transfer from the a-c grid-potential source to the electrons, resulting in an increase in input conductance.

Ferris showed that magnitude of the input admittance Y and input conductance g_1 of a triode tube may be represented approximately by

$$Y = k_1 g_m f T \tag{1}$$

$$g_1 = k_2 g_m f^2 T^2 \tag{2}$$

where g_m is the transconductance of the tube, f is the frequency, T is a quantity which is proportional to the electron transit time, and k_1 and k_2 are proportionality constants. Equations (1) and (2) are similar to (4.03-10).

In the preceding discussion, we were concerned with energy transfer from the a-c grid-potential source to the electrons. Let us now consider the energy transfer with respect to the a-c plate potential. Ideally, the electrons should flow through the grid-plate region during the retarding phase of the a-c plate potential so that energy is transferred from the electrons to the source of a-c plate potential, *i.e.*, the load impedance. How-

ever, for large electron transit angles, the a-c plate potential may reverse its phase before the electrons arrive at the plate. Energy is then transferred from the source of the a-c plate potential (the load impedance) to the electrons during a portion of the cycle, thereby decreasing the power output and efficiency.

We therefore conclude that electron transit-time delay may (1) increase the input conductance and (2) decrease the power output and efficiency of the tube.

The cathode lead inductance and input capacitance have an important bearing upon the input admittance of the tube. These form a series *L-C*
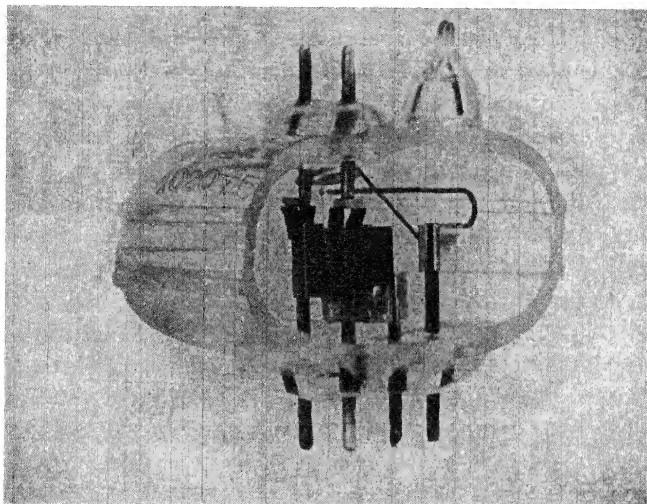
FIG. 1.—W.E. 368A triode. (*Courtesy of the Western Electric Company.*)

circuit, the admittance of which increases as the frequency approaches the resonant frequency. As a typical example, the 955 acorn tube has an input capacitance of 1 micromicrofarad and lead inductance of approximately 0.015 microhenry. The resonant frequency of the input circuit, taken alone, is approximately 1,330 megacycles. Obviously the other circuit parameters influence the value of the resonant frequency; however, this example serves to indicate one of the serious limitations in the operation of conventional types of tubes at microwave frequencies. It can be shown that the input conductance increases approximately as the square of the frequency, hence its variation with frequency is similar to the transit-time conductance of Eq. (5.01-2). In many types of tubes, the lead-inductance effects offer a more serious limitation than transit-time effects.

Grounded-grid circuits are frequently used in amplifiers and oscillators at frequencies above 100 megacycles. Grounding the grid serves to reduce

the input conductance due to transit-time effects and lead inductance. Further advantage is gained in amplifiers due to the fact that grounding the grid tends to shield the input circuit from the output circuit. This shielding is similar to that resulting from the use of the screen grid in the pentode. The shielding reduces the tendency toward oscillation and makes it possible to obtain increased gain.

5.02. Triode Tubes and Oscillator Circuits.—Several triode tubes, designed for operation at microwave frequencies, are shown in Figs. 1, 2, and 3. The W.E. 368A tube of Fig. 1 has a cylindrical cathode, grid, and anode. The triangular carbon block attached to the anode increases the heat-dissipation area. Leads are brought out at both ends of the tube to facilitate the use of two tuned circuits, thereby dividing the interelectrode capacitance between the two tuned circuits. This tube has an upper frequency limit of approximately 1,700 megacycles and can deliver a power output of 10 watts at 500 megacycles.

FIG. 2.—G.E. 2C40 lighthouse tube. (*Courtesy of the General Electric Company.*)

Figure 2 shows the G.E. 2C40 lighthouse triode in which the cathode, grid, and anode form parallel planes. The cathode is indirectly heated and
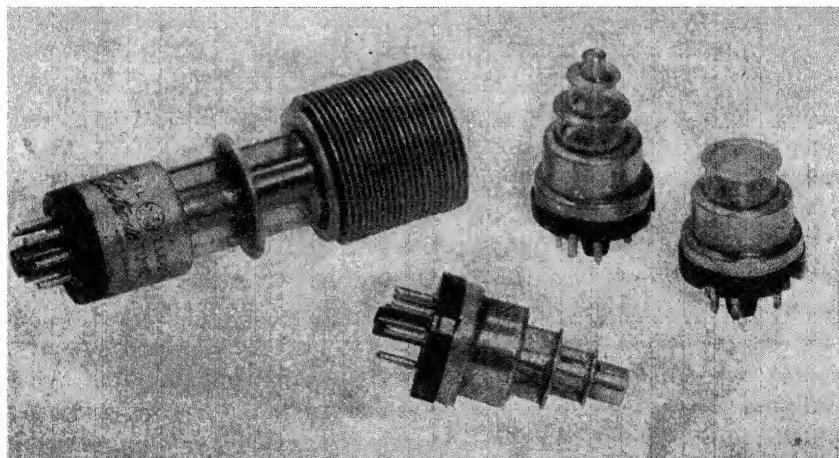
FIG. 3.—Lighthouse tubes. (*Courtesy of the General Electric Company.*)

has an activated oxide-emitting layer on the flat end of the cylinder. The grid consists of a woven tungsten mesh which is brazed to an annular ring. The plate is a single steel cylinder machined from solid stock and silver plated. This tube is designed for use in coaxial resonators, such as those shown in Fig. 6. The lighthouse tube may be used either as an amplifier or as an oscillator at frequencies as high as 3,500 megacycles.
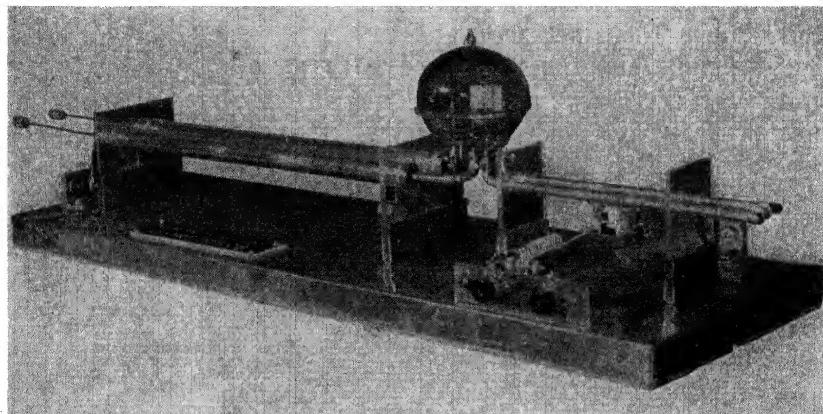
(a)
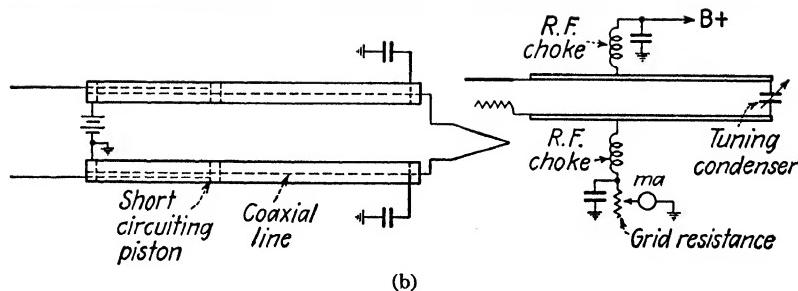
(b)

FIG. 4.—Triode oscillator using resonant lines.

Figure 3 shows several modifications of lighthouse tubes including the GL522 transmitter tube which has a power output rating of 25 watts at a frequency of 500 megacycles with a plate voltage of 1,000 volts.

At frequencies above approximately 200 megacycles, the resonant systems used in oscillators and amplifiers are usually constructed of low-loss short-circuited or open-circuited transmission lines, or of lines which are terminated in pure reactances. The properties of such lines in the vicinity of the antiresonant frequency are similar to those of a parallel *L-C* circuit. However, it is possible to obtain much higher effective *Q*'s and more stable

performance using transmission lines as resonant circuits than is possible with lumped circuits. The frequency of oscillation is approximately the frequency at which the system is antiresonant. In Sec. 10.03 it is shown that a line which is either open-circuited or short-circuited at the distant end and which is shunted by a capacitance at the input terminals, *i.e.*, the interelectrode capacitance of the tube, will be antiresonant when

$$\tan \frac{\omega l}{v_c} = \frac{1}{\omega C Z_0} \qquad \text{short-circuited line} \qquad (1)$$

$$\tan \frac{\omega l}{v_c} = -Z_0 \omega C \qquad \text{open-circuited line} \qquad (2)$$

where l and Z_0 are the length and characteristic impedance of the line, respectively, C is the shunt capacitance, ω is the angular frequency, and
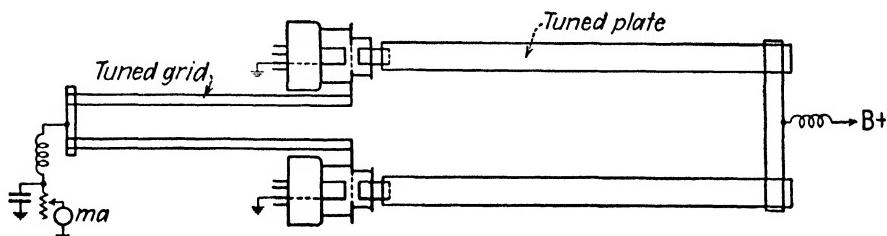
FIG. 5.—Push-pull triode oscillator using lighthouse tubes.

v_c is the velocity of light. Equation (1) or (2) may be used to determine the length of line required for a given frequency of oscillation.

Figure 4 shows an oscillator using a W.E. 316A "doorknob" tube. The tuned circuit consists of an open-wire transmission line in the grid-plate circuit which is tuned by means of a small condenser at the distant end of the line. The d-c connections to the grid and plate are made at the approximate radio-frequency voltage nodes on the grid-plate line. Chokes are provided to prevent radio-frequency power loss at these junctions. The filament circuit contains a pair of coaxial lines which are tuned by means of short-circuiting pistons. The transmission lines serve as chokes to isolate the cathode from ground and also to have the d-c filament connections at radio-frequency voltage nodal points on the line. This oscillator will deliver 6 to 8 watts of power at an efficiency of 40 per cent at frequencies up to 600 megacycles. The load is coupled to the grid-plate line.

A push-pull oscillator is shown in Fig. 5. The tuned circuits consist of short-circuited lines in the grid and plate circuits. Since the short-circuited ends of the lines correspond to voltage nodes, the d-c connections are made at these points.

Two cavity oscillators using lighthouse tubes are shown in Fig. 6. The cavity of Fig. 6a contains short-circuited coaxial lines in the cathode-grid and grid-plate circuits. These are tuned by means of sliding pistons which are provided with spring contact fingers to make electrical contact with the cylindrical walls of the resonator. The cavity of Fig. 6a may be used as an amplifier by feeding the input signal into the cathode-grid resonant circuit by means of a coupling loop or probe. Power output is obtained
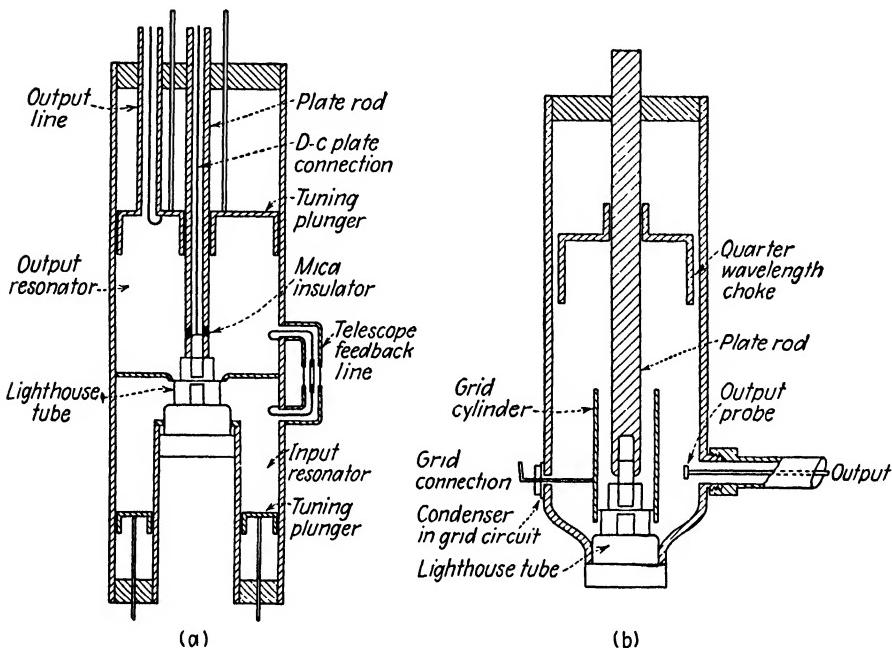
Fig. 6.—Cavity oscillators using lighthouse tubes.

through a similar coupling loop or probe in the grid-plate line. When this unit is used as an oscillator the coupling loops in the input and output circuits are connected together to provide feedback.

The resonant cavity shown in Fig. 6b contains an open-ended grid cylinder, a cylindrical plate rod, and a choke consisting of a quarter-wavelength section of open-circuited line. The grid cylinder and plate rod form the resonant line of the grid-plate circuit, whereas the grid cylinder and outer shell of the cavity form the resonant line in the cathode-grid circuit. Since the grid cylinder is open at one end, the input and output circuits are coupled together, thereby forming a reentrant oscillator. The quarter-wavelength choke prevents energy from escaping out of the back end of the cavity. Tuning is accomplished by sliding the plate rod on the plate cap of the lighthouse tube. The end of the plate rod has spring contact

fingers which firmly grip the plate cap of the tube, at the same time permitting a small amount of longitudinal motion of the plate rod. This motion of the plate rod has the effect of changing the characteristic impedance over a small portion of the grid-plate coaxial line (due to the difference in diameter of the plate rod and plate cap of the tube) and this, in turn, alters the resonant frequency of the line. It is also necessary to adjust the position of the choke as the cavity is tuned. This may be accomplished by means of separate adjustments of the plate rod and choke or, with proper design, the choke may be attached to the plate rod so as to provide a single-control tuning mechanism.

With lighthouse-tube oscillators of the type shown in Fig. 6, it is possible to obtain continuous power outputs of several watts at wavelengths as low as 8 centimeters or frequencies as high as 3,800 megacycles.

5.03. Criterion of Oscillation.—A negative-grid oscillator consists essentially of an amplifier in which a small fraction of the output voltage

FIG. 7.—Block diagram of an oscillator.

is fed back into the input in order to sustain oscillation. In order for oscillation to exist, certain requirements must be satisfied. The criterion of oscillation enables us to determine the conditions of oscillation as well as the oscillating frequency.

Consider the block diagram shown in Fig. 7. This represents an amplifier having a voltage gain K, and a feedback circuit having a complex voltage ratio β. In the following analysis class A operation is assumed. Transit-time effects are assumed to be negligible.

Let $\mathbf{E}_g$ and $\mathbf{E}_p$ represent the input and output voltages of the amplifier, respectively. The feedback voltage is $\beta\mathbf{E}_p$, where the value of β is determined by the feedback circuit. We then have

$$\mathbf{E}_g = \beta\mathbf{E}_p \qquad (1)$$

Also, the voltage gain is defined by

$$K = \frac{\mathbf{E}_p}{\mathbf{E}_g} \qquad (2)$$

Combining Eqs. (1) and (2), we obtain the criterion of oscillation

$$\beta K = 1 \qquad (3)$$

Let us now apply this criterion to the general type of oscillator shown in Fig. 8. This circuit may be used to represent any one of the oscillators shown in Figs. 4 to 6. Thus, comparing Figs. 6a and 8, the impedance Z_1 in Fig. 8 represents the resonant line in the cathode-grid circuit of Fig. 6a,
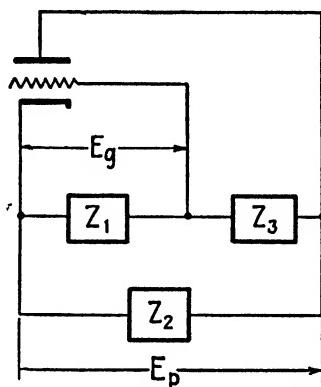
FIG. 8.—Block diagram representing the oscillators shown in Figs. 4 and 6.

Z_3 represents the line in the grid-plate circuit, and Z_2 is the cathode-plate capacitance and the effect of the coupling loop.

From Fig. 8, we obtain

$$\beta = \frac{E_g}{E_p} = \frac{Z_1}{Z_1 + Z_3} \tag{4}$$

The gain of the stage is

$$K = -\frac{\mu Z_L}{r_p + Z_L} = -\frac{\mu}{r_p Y_L + 1} \tag{5}$$

where $Z_L = 1/Y_L$ is the load impedance. Again, from Fig. 8, we obtain

$$Y_L = \frac{Z_1 + Z_2 + Z_3}{Z_2(Z_1 + Z_3)} \tag{6}$$

Inserting this expression for Y_L into Eq. (5) yields

$$K = \frac{-\mu(Z_1 + Z_3)Z_2}{r_p(Z_1 + Z_2 + Z_3) + Z_2(Z_1 + Z_3)} \tag{7}$$

Substituting β from Eq. (4) and K from Eq. (7) into (3), the criterion of oscillation yields

$$\frac{-\mu Z_1 Z_2}{r_p(Z_1 + Z_2 + Z_3) + Z_2(Z_1 + Z_3)} = 1 \tag{8}$$

Let us now assume that the impedances are pure reactances,[1] *i.e.*, $Z_1 = jX_1$, $Z_2 = jX_2$, and $Z_3 = jX_3$. By making these substitutions in Eq. (8), we obtain

$$-X_2[X_1(1 + \mu) + X_3] + jr_p(X_1 + X_2 + X_3) = 0 \qquad (9)$$

Equating imaginary terms to zero, we have

$$X_1 + X_2 + X_3 = 0 \qquad (10)$$

Likewise, equating the real terms to zero and substituting $X_1 + X_3 = -X_2$ from Eq. (10), we obtain

$$\mu = -\frac{(X_1 + X_3)}{X_1} = \frac{X_2}{X_1} \qquad (11)$$

Equation (11) expresses a critical value of amplification factor which is required for oscillation. In the actual performance of an oscillator, the amplification factor of the tube is not constant but, rather, varies over a relatively wide range of values. Consequently, Eq. (11) may be interpreted as representing somewhat of an average value of μ required for oscillation. Equation (10) determines the frequency of oscillation. To satisfy this relationship and also obtain a positive value of μ in Eq. (11), it is necessary that X_1 and X_2 be similar types of reactance, whereas X_3 must be a reactance of the opposite type.

In the oscillator of Fig. 6a, the reactances X_1 and X_3 each consist of the input reactance of a short-circuited line shunted by the interelectrode capacitance of the tube. Since the expressions for these reactances are quite long, a complete analysis will not be attempted. However, in order to illustrate the method, let us assume that X_1 and X_2 are both capacitive with effective capacitances C_1 and C_2, respectively, and that X_3 is inductive with an effective inductance L_3. Inserting these into Eq. (11), we obtain the value of μ

$$\mu = \frac{C_1}{C_2}$$

Inserting the reactances into Eq. (10), we obtain the angular frequency of oscillation.

$$\omega L_3 - \frac{1}{\omega C_1} - \frac{1}{\omega C_2} = 0$$

$$\omega = \sqrt{\frac{1}{L_3}\left(\frac{1}{C_1} + \frac{1}{C_2}\right)} \qquad (12)$$

---

[1] This assumption corresponds to assuming a lossless system. It will be shown later that the input impedance of a lossless short-circuited line is a pure reactance except at the resonant and antiresonant frequencies where it has zero and infinite values, respectively.

5.04. Analysis of the Class C Oscillator.—Let us briefly consider the factors involved in class C oscillator performance, ignoring electron transit-time effects.

For high-efficiency operation, the d-c grid-bias voltage should be approximately one and one-half to three times the cutoff value. Oscillators are usually self-biased, the bias voltage resulting from the rectified grid current flowing through a resistance in the grid circuit. Cathode bias is also pro-
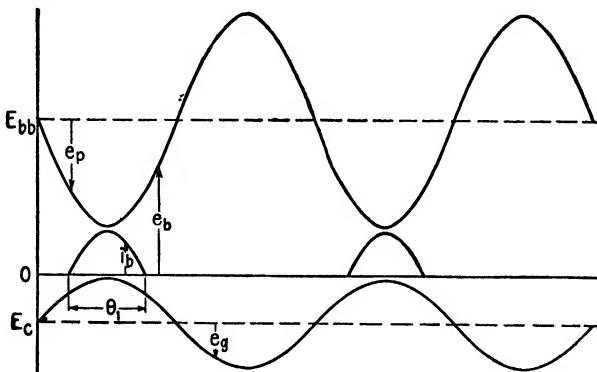
Fig. 9.—Voltage and current relationships in a class C oscillator.

vided in some cases. In a well-designed oscillator, the maximum grid voltage $e_{c\,max}$ is approximately equal to the minimum plate voltage $e_{b\,min}$, that is,

$$e_{c\,max} \approx e_{b\,min} \tag{1}$$

The rms alternating plate voltage $\mathbf{E}_{p1}$ is

$$\mathbf{E}_{p1} = \frac{\mathbf{E}_{bb} - e_{b\,min}}{\sqrt{2}} \tag{2}$$

where $\mathbf{E}_{bb}$ is the d-c plate-supply voltage.

Plate current flows only during a brief conduction angle as shown in Fig. 9. The plate current pulses may be analyzed by Fourier series to obtain the average value of plate current I_{b0} and the rms fundamental component I_{p1} in terms of the maximum plate current $i_{b\,max}$. Assume that the plate current has the same waveform as the grid voltage during the conduction portion of the cycle. If the total conduction angle as shown in Fig. 9 is θ_1, this analysis yields the relationships [1]

$$I_{b0} = \frac{i_{b\,max}}{\pi(1 - \cos\theta_1/2)}(\sin\theta_1/2 - \theta_1/2\cos\theta_1/2) \tag{3}$$

$$I_{p1} = \frac{i_{b\,max}}{\sqrt{2}\pi(1 - \cos\theta_1/2)}\left(\theta_1/2 - \frac{\sin\theta_1}{2}\right) \tag{4}$$

[1] EVERITT, W. L., "Communication Engineering," 2d ed., pp. 565–577, McGraw-Hill Book Company, Inc., New York, 1937.

Curves of $I_b/i_{b\,max}$ and $I_{p1}/i_{b\,max}$ against θ_1 are shown in Fig. 10.

For high-efficiency operation, the total conduction angle is approximately $\theta_1 = 125$ degrees to 150 degrees, yielding values of $I_b/i_{b\,max} = 0.2$ to 0.24 and $I_{p1}/i_{b\,max} = 0.25$ to 0.28.

Assume that the frequency of oscillation is equal to the antiresonant frequency of the load impedance (taken from cathode to plate). The load is then a pure resistive impedance which we represent by R_L. The alter-
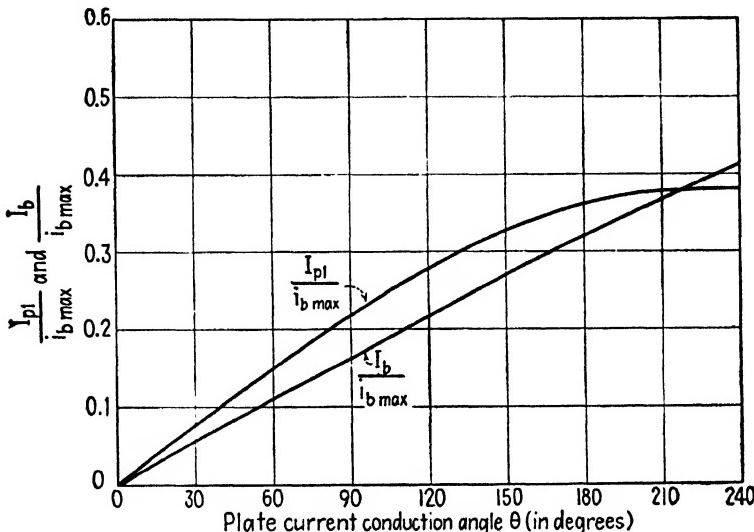
FIG. 10.—Curves of $\dfrac{I_b}{i_{b\,max}}$ and $\dfrac{I_{p1}}{i_{b\,max}}$ plotted against plate-current conduction angle θ.

nating voltage developed across R_L is $\mathbf{E}_{p1} = I_{p1}R_L$. Inserting $\mathbf{E}_{p1}$ from Eq. (2) into this relationship, we obtain

$$R_L = \frac{\mathbf{E}_{p1}}{I_{p1}} = \frac{\mathbf{E}_{bb} - e_{b\,min}}{\sqrt{2}I_{p1}} \tag{5}$$

Let P_g be the grid driving power. The a-c power output P_{ac}, d-c power input P_b, plate dissipation P_p, and plate-circuit efficiency η_p are then

$$P_{ac} = \mathbf{E}_{p1}I_{p1} - P_g \tag{6}$$

$$P_b = \mathbf{E}_{bb}I_{b0} \tag{7}$$

$$P_p = P_b - \mathbf{E}_{p1}I_{p1} \tag{8}$$

$$\eta_p = \frac{P_{ac}}{P_b} = \frac{\mathbf{E}_{p1}I_{p1} - P_g}{\mathbf{E}_{bb}I_{b0}} \tag{9}$$

From an electronic viewpoint P_b represents the power transferred from the d-c plate potential source to the electron stream. The power $\mathbf{E}_{p1}I_{p1}$

is the power transferred from the electron stream to the load impedance as a result of retardation of the electrons by the a-c component of field intensity. A portion P_g of this power output is returned to the grid circuit to sustain oscillation. The grid driving power is given by the approximate expression $P_g = \mathbf{E}_g I_{g0}$, where $\mathbf{E}_g$ is the maximum value of a-c grid voltage and I_{g0} is the d-c grid current.

The load impedance has an important bearing upon the power output and efficiency of the oscillator, since it influences the values of $\mathbf{E}_{p1}$ and I_{p1}. The desired value of load impedance may be computed from Eq. (5). However, in practice, it is usually difficult to calculate the impedance of resonant lines with coupled loads. Consequently the load impedance is usually adjusted experimentally by varying the coupling between the load and the resonant line until maximum power output is obtained.

At microwave frequencies, the electron transit-time delay increases the input conductance of the tube, thereby resulting in an increase in grid driving power. Electron transit-time delay also introduces a phase shift between the maximum plate current and the minimum plate voltage. These factors, together with the increased losses at microwave frequencies, result in reduced power output and lower efficiency.

5.05. Frequency Stability of Triode Oscillators.—The conditions required for good frequency stability of oscillators are somewhat different from those required for high power output. Good frequency stability necessitates high-Q circuits which, in turn, implies low power output. The feedback circuits and load circuits should be loosely coupled to the output circuit of the oscillator in order to minimize the loading on this circuit.

For good frequency stability, the resonant circuit should present a relatively high impedance to the tube at the frequency of oscillation. The interelectrode capacitance of the tube has the effect of decreasing the length of line required for a given frequency of oscillation and also of reducing the antiresonant input impedance. In general, this effect is less for coaxial lines than for open-wire lines because of the fact that coaxial lines have a lower L/C ratio. Coaxial lines are therefore preferred in microwave systems. Coaxial lines are also self-shielding, thereby tending to minimize the radiation losses from the line.

Another important factor in the frequency stability of an oscillator is the thermal expansion of the tube elements and the associated circuits, with changes in temperature. For example, a resonant line constructed of brass would have a temperature coefficient of expansion of 19 parts per million per degree centigrade. Tests on an oscillator of the type shown in Fig. 6b over a temperature range of 140 degrees centigrade have shown an average frequency drift of 20 parts per million per degree centigrade. This shows a striking correlation between the thermal expansion and the frequency drift.

Various temperature-compensating devices have been developed to compensate for thermal expansion. One such device has a small air condenser at one end of the line which is arranged so that expansion or contraction of the line changes the spacing between the condenser plates in such a manner as to maintain constant oscillator frequency.[1]

According to the criterion of oscillation, it is possible for the frequency to differ slightly from the antiresonant frequency of the output circuit. This occurs when the phase shift in the feedback circuit is not exactly 180 degrees. If electron transit-time effects are negligible, maximum frequency stability occurs when the phase shift in the feedback circuit is exactly 180 degrees. In microwave oscillators, the phase shift in the feedback circuit should be slightly greater than 180 degrees in order to compensate for transit-time delay.

With careful oscillator design, using high-Q circuits and a well-regulated power supply, it is possible to achieve a frequency stability of the order of 10 to 20 parts per million, which is somewhat poorer than the frequency stability attainable with crystal oscillators at lower frequencies.

5.06. Amplifiers Using Negative-grid Triodes.—Most triode tubes which are designed for use at microwave frequencies may be used either as amplifiers or as oscillators. The lighthouse tube and cavity shown in Fig. 6a may be used as an amplifier by disconnecting the feedback cable and feeding the input signal into the cathode-grid region of the cavity. The power output is taken from the grid-plate region of the cavity.

The power gain of amplifiers operating at microwave frequencies is relatively low, being of the order of from 5 to 20. Since the bandwidth of tuned circuits at these frequencies is relatively large, even for high-Q circuits, the tuned circuits admit a relatively large amount of noise. This results in a low signal-to-noise ratio which imposes a serious limitation upon the use of amplifiers in amplitude-modulated transmitters and receivers at microwave frequencies. If frequency modulation is used, the noise may be largely separated from the signal in the limiter stage at the receiver; hence the presence of noise in the amplifier is not so serious a limitation in frequency modulation as in amplitude modulation.

[1] HANSELL, C. W., and P. S. CARTER, Frequency Control by Low Power Factor Line Circuits, *Proc. I.R.E.*, vol. 24, pp. 597–619; April, 1936.

CHAPTER 6

TRANSIT-TIME OSCILLATORS

It has been shown that detrimental effects are encountered in the nega-tive-grid triode oscillator or amplifier if the electron transit time exceeds a small fraction of the period of the alternating cycle. There are other types of oscillators, however, in which operation is dependent upon a defi-nite relationship between the electron transit time and the period of the alternating cycle. These transit-time oscillators include positive-grid oscil-lators, klystrons, resnatrons, traveling-wave tubes, and magnetrons.

POSITIVE-GRID OSCILLATOR

6.01. Operation of the Positive-grid Oscillator.—The positive-grid os-cillator consists of either a parallel-plane or a cylindrical-element triode tube with the grid at a positive d-c potential with respect to its cathode and plate as shown in Figs. 1 and 2. The tuned circuit may consist of either a lumped *L-C* circuit or a distributed parameter system such as a short-circuited line, an open-circuited line, or a resonant cavity. In the opera-tion of the positive-grid oscillator, an electron space-charge cloud oscillates back and forth about a mean position corresponding to the grid plane. The period of electron oscillation is determined by the tube geometry and the electrode potentials. The oscillating space charge induces an alter-nating component of current in the external circuit and the resulting voltage drop across the load impedance produces an alternating field in the interelectrode space of the tube. A majority of the electrons oscillate in such a phase as to be retarded by the alternating field, and hence transfer a portion of their energy to the resonant circuit during each cycle of oscillation.

Let us assume, for the moment, that the grid of the tube shown in Fig. 1 has a positive d-c applied potential and that the load impedance is removed from the circuit. Consider the motion of an electron in the resulting d-c field. Since the grid is positive with respect to both the cathode and plate, the electric field is in such a direction as to accelerate the electron when it travels toward the grid plane and decelerate the electron when it travels away from the grid plane.

The electron is emitted from the cathode and is accelerated as it ap-proaches the grid plane. Passing between the grid wires, the electron enters the grid-plate region where it is decelerated. It comes momentarily

to rest in the vicinity of the plate, reverses its direction of travel, and is accelerated during its return journey to the grid plane. If the electron again passes between the grid wires, it is decelerated as it approaches the cathode, reverses its direction of travel in the vicinity of the cathode, and once again starts back toward the grid. The electron continues to oscillate back and forth with approximately equal amplitudes on either side of the grid plane in a manner similar to the oscillation of a frictionless pendulum, until it eventually collides with a grid wire.

When the load impedance is connected as shown in Fig. 1 and oscillating conditions prevail, the a-c potential developed across the load impedance
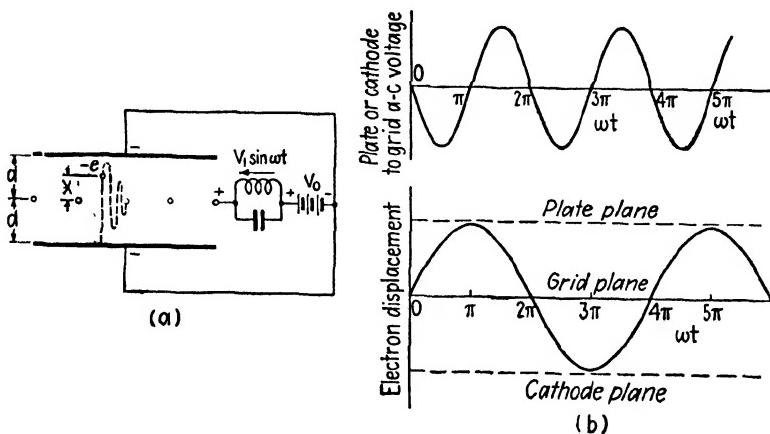
FIG. 1.—Positive-grid oscillator.

produces an alternating field superimposed upon the d-c field in the inter-electrode space. In the oscillator of Fig. 1, the frequency of the alternating potential is twice the frequency of electron oscillation. The phase relationships between the electron displacement and the alternating potential for the most favorable electron are shown in Fig. 1b. For this particular electron, the alternating component of grid potential (with respect to cathode or plate) is negative as the electron approaches the grid from either direction and positive as the electron recedes from the grid. Consequently, this particular electron will be retarded by the a-c field throughout its entire cycle of oscillation and will transfer energy to the external tuned circuit. The kinetic energy and the amplitude of oscillation of the electron both decrease with each successive cycle of electron oscillation. In this respect, oscillation of the electron is analogous to the swing of a damped pendulum.[1]

---

[1] The oscillation of the electron is not entirely analogous to the oscillation of a damped pendulum, since the retarding force of the alternating field does not obey the same laws as the frictional force of the pendulum.

The phase relationships shown in Fig. 1b are those of an ideal electron, *i.e.*, an electron which transfers energy to the external circuit throughout its entire cycle of oscillation. We shall see presently that the period of oscillation of an electron changes continuously during successive cycles of oscillation. Consequently, an electron which starts to oscillate in the ideal phase will gradually fall out of phase with the alternating potential, thereby oscillating in a less favorable phase. Since electrons are emitted from the cathode at a more or less uniform rate, part of the electrons will depart from the cathode in such phase as to take energy from the alternating field. By virtue of their increased kinetic energy, these electrons travel all the way to the plate or cathode within a few cycles of oscillation and therefore are withdrawn from active duty. Hence we find that those electrons which, on the average, give energy to the alternating field remain in oscillation, whereas those electrons which take energy from the alternating field travel to either the plate or cathode and are thereby excluded.

6.02. Analysis of the Positive-grid Oscillator.—To illustrate the analytical relationships involved in positive-grid oscillators, consider the triode tube shown in Fig. 1. The grid is midway between the cathode and anode. It is assumed that the space-charge density is small and that it does not affect the electric field distribution in the interelectrode space. Distances are measured from the grid plane, positive values of x being taken in the direction of the plate.

Consider now an electron moving in the grid-plate region. The potential difference between grid and plate is taken as $V = -V_0 - V_1 \sin \omega t$, the negative sign of the d-c potential signifying that the grid is positive with respect to the plate. Equation (2.07-8) may be used to express the electron displacement in the grid-plate region if the potentials V_0 and V_1 in this equation are replaced by $-V_0$ and $-V_1$. The electron displacement equation then becomes

$$x = \frac{k(\omega T)^2}{2} + \frac{kV_1}{V_0}[(\omega T - \sin \omega T)\cos \phi + (1 - \cos \omega T)\sin \phi]$$
$$+ \frac{v_0(\omega T)}{\omega} \quad (1)$$

In this equation ωT is the electron transit angle; *i.e.*, the phase angle through which the alternating voltage varies while the electron is in flight from the grid to a point distant x from the grid; ϕ is the phase angle of the alternating potential at the instant when the electron leaves the grid; and k is given by $k = -eV_0/\omega^2 md$.

We now assume that the electron completes one half cycle of oscillation and returns to the grid while the alternating potential completes one full cycle. Hence the electron leaves the grid plane ($x = 0$) when $\omega T = 0$ with an initial velocity v_0 and returns to the grid plane when $\omega T = 2\pi$. Writing

Eq. (1) for the instant at which the electron returns to the grid plane by setting $x = 0$ and $\omega T = 2\pi$, we obtain

$$\pi k + \frac{kV_1}{V_0} \cos \phi + \frac{v_0}{\omega} = 0 \tag{2}$$

Substituting $k = -eV_0/\omega^2 md$ and solving for ω, we obtain the angular frequency of electrical oscillation,

$$\omega = \frac{eV_0}{v_0 md} \left(\pi + \frac{V_1}{V_0} \cos \phi \right) \tag{3}$$

The period of electron oscillation T_1 was assumed to be twice the period of electrical oscillation, or $T_1 = 2/f = 4\pi/\omega$. Substitution of Eq. (3) for ω yields the period of electron oscillation

$$T_1 = \frac{4\pi v_0 md}{eV_0[\pi + (V_1/V_0) \cos \phi]} \tag{4}$$

Equation (4) shows that the period of electron oscillation is dependent upon the initial phase angle ϕ. The ideal phase is that shown in Fig. 1b, corresponding to $\phi = 0$ radians. Since electrons cross the grid plane at various values of ϕ, the period (or frequency) of electron oscillation will differ slightly for various electrons, each according to its particular value of ϕ. Furthermore, Eq. (4) shows that the period of electron oscillation varies directly with the electron velocity v_0 at the grid plane. If an electron oscillates in such a phase as to give energy to the alternating field, its kinetic energy and velocity v_0 at the grid plane decrease for each successive cycle of oscillation and, according to Eq. (4), the period of electron oscillation decreases or its frequency increases. This shift from a favorable phase of oscillation to an unfavorable phase results in an appreciable reduction in power output and efficiency.

Instead of finding a simple picture in which all electrons have identical periods of oscillation, we have a much more complicated situation in which the frequency of electron oscillation differs for various electrons and undergoes continuous change for any one electron. The frequency of electrical oscillation is determined by the composite effects of all the electrons oscillating in the interelectrode space, its value being approximately twice the average frequency of electron oscillation.

The velocity v_0 in Eq. (4) is determined largely by the d-c potential; hence we may assume that $v_0 = \sqrt{2V_0e/m}$. The period of electron oscillation then becomes

$$T_1 = \frac{4\pi d \sqrt{2m/eV_0}}{[\pi + (V_1/V_0) \cos \phi]} \tag{5}$$

Since the a-c potential is usually much smaller than the d-c potential, we have $V_1/V_0 \ll \pi$ and, to a first approximation, the term $(V_1/V_0) \cos \phi$ in Eq. (5) may be neglected. The frequency of electrical oscillation is $f = 2/T_1$ and the corresponding wavelength is given by $\lambda = v_c/f$, where $v_c = 3 \times 10^8$ meters per second is the velocity of light. The period of
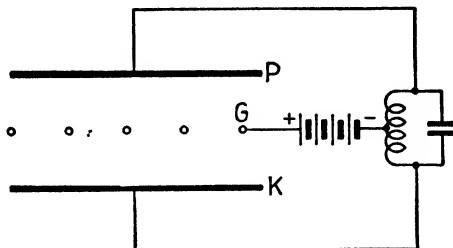
FIG. 2.—Positive-grid oscillator with tuned circuit connected between cathode and plate.

electron oscillation, frequency of electrical oscillation, and wavelength then become approximately

$$T_1 = 4d\sqrt{\frac{2m}{eV_0}} = 1.35 \times 10^{-5}\,\frac{d}{\sqrt{V_0}} \qquad \text{seconds} \qquad (6)$$

$$f = \frac{1}{d}\sqrt{\frac{eV_0}{8m}} = \frac{1.48 \times 10^5}{d}\sqrt{V_0} \qquad \text{cycles per second} \qquad (7)$$

$$\lambda = v_c d\sqrt{\frac{8m}{eV_0}} = 2020\,\frac{d}{\sqrt{V_0}} \qquad \text{meters} \qquad (8)$$

In these equations d is the cathode-to-grid and grid-to-plate distance in meters and V_0 is in volts. Comparison of Eqs. (6) and (2.07-11) shows that the period of electron oscillation is approximately four times the time required for the electron to travel from the cathode to the grid in a d-c field.

In another form of the positive-grid oscillator, the resonant circuit is connected between cathode and plate as shown in Fig. 2. With this arrangement, the frequency of electrical oscillation is approximately equal to the average frequency of electron oscillation. Consequently the electrical frequency will be one-half that given by Eq. (7) and the wavelength will be double the value given by Eq. (8). The wavelength for this type of oscillation is, therefore,

$$\lambda = 4040\,\frac{d}{\sqrt{V_0}} \qquad \text{meters} \qquad (9)$$

The wavelength of oscillation for the cylindrical-electrode positive-grid oscillator [1] with cathode and anode at the same potential and tuned circuit connected between cathode and anode, similar to that shown for the parallel-plane triode of Fig. 2, is

$$\lambda = \frac{2000 r_g}{\sqrt{V_0}} \left[f(x) + g(y) \right] \qquad \text{meters} \qquad (10)$$
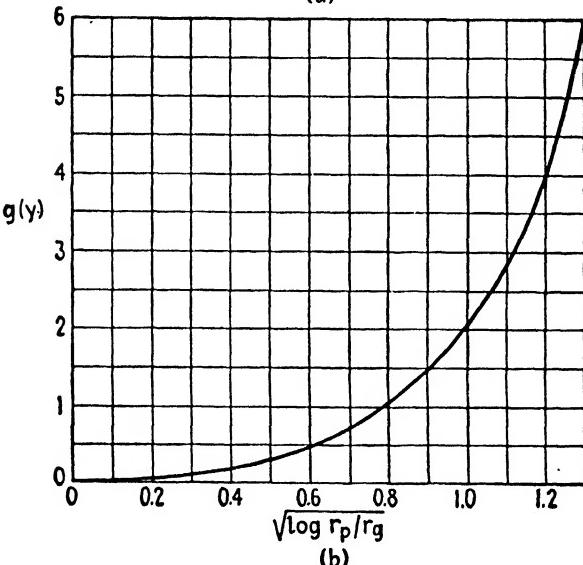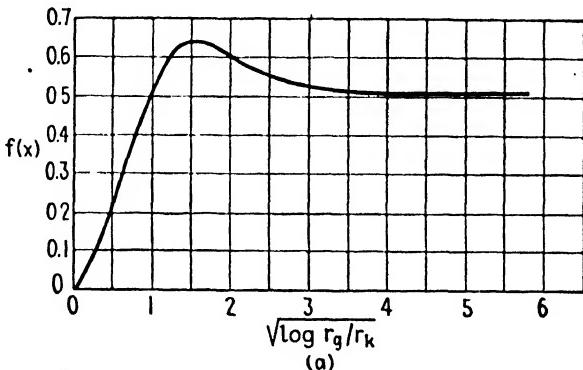
Fig. 3.—Plot of $f(x)$ and $g(y)$ for the cylindrical-electrode positive-grid oscillator.

where V_0 is the grid-cathode d-c potential. The quantities $f(x)$ and $g(y)$ are functions of the cathode radius r_k, grid radius r_g, and plate radius r_p. These functions are plotted in Fig. 3.

[1] SCHEIBE, A., The Generation of Ultra Short Waves with Hot Cathode Tubes *Ann. Physik*, vol. 73, no. 54, pp. 54–88; 1924.

6.03. Operation of the Positive-grid Oscillator.—Positive-grid oscillations may exist even though the resonant frequency of the external circuit differs from the fundamental frequency or a harmonic of the frequency of electron oscillation. Under these conditions, the external circuit has very little effect upon the frequency of electrical oscillation, this being determined by the frequency of electron oscillation. Oscillations of this type are known as Barkhausen-Kurz oscillations.
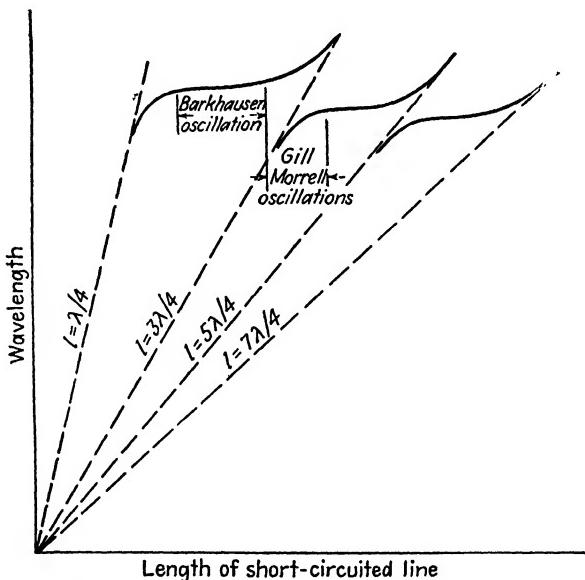
Fig. 4.—Effect of tuning of the external circuit upon wavelength of a positive-grid oscillator.

An abrupt change in wavelength and a pronounced increase in alternating voltage and power output occur when the external circuit is tuned to the fundamental or a harmonic of the frequency of electron oscillation. Oscillations of this variety are known as Gill-Morrell oscillations. The change in wavelength of a positive-grid oscillator with tuning of the external circuit is shown in Fig. 4.

Positive-grid oscillators, in general, have relatively low power output and efficiency. This is due, in part, to the fact that a high percentage of the electrons do not oscillate in the most favorable phase and also due to the tendency of electrons to shift from a favorable phase to an unfavorable phase as they continue to oscillate. Since most of the electrons terminate at the grid, the grid losses are considerably greater in the positive-grid oscillator than in negative-grid oscillators. Consequently, the grid must be capable of radiating a relatively large amount of heat.

The period of electron oscillation is dependent upon the d-c voltage. A carefully regulated power supply is therefore required for high frequency

stability.   The positive-grid oscillator may be amplitude modulated by applying a relatively small modulating voltage in series with the d-c supply voltage.   Due to the inherently poor frequency stability, however, there is a considerable amount of frequency modulation along with the amplitude modulation.

KLYSTRON OSCILLATOR

6.04. Description of the Klystron Oscillator.—In the negative-grid triode oscillator, the electrons travel through superimposed d-c and a-c
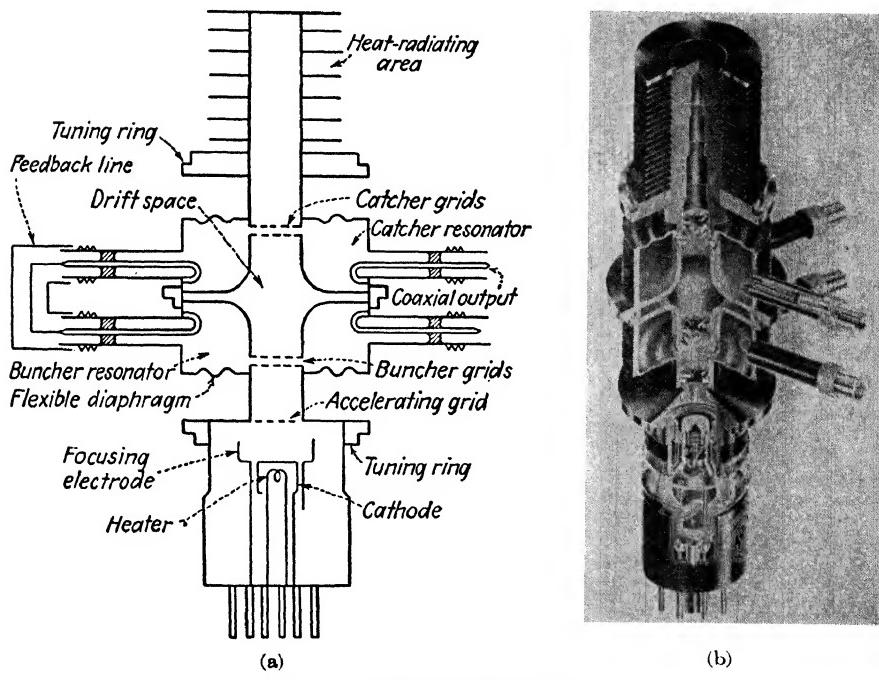
FIG. 5.—Double-resonator klystron.

electric fields.   Electrons are simultaneously accelerated by the d-c field and retarded by the a-c field.   If the retarding a-c field is comparable with the d-c field, the electrons never attain a very high velocity and therefore the electron transit angle may be large at microwave frequencies.   Several types of tubes have been developed in which the electrons are initially accelerated to a high velocity by a d-c field before they enter the retarding a-c field.   The electrons then have a relatively high velocity as they travel through the a-c field; hence the electron transit angle is greatly reduced. The double-resonator klystron and the reflex klystron are examples of such tubes.

The double-resonator klystron, shown in Figs. 5 and 6, contains an oxide-coated cathode, a control grid, two metallic resonators, and a collector anode. The resonator nearest the cathode is known as the *buncher* or *input resonator* and the second resonator is the *catcher* or *output resonator*. The entire assembly is evacuated and the high-frequency electrical connections consist of small coupling loops placed inside the resonators with vacuum-sealed coaxial fittings for external connections.
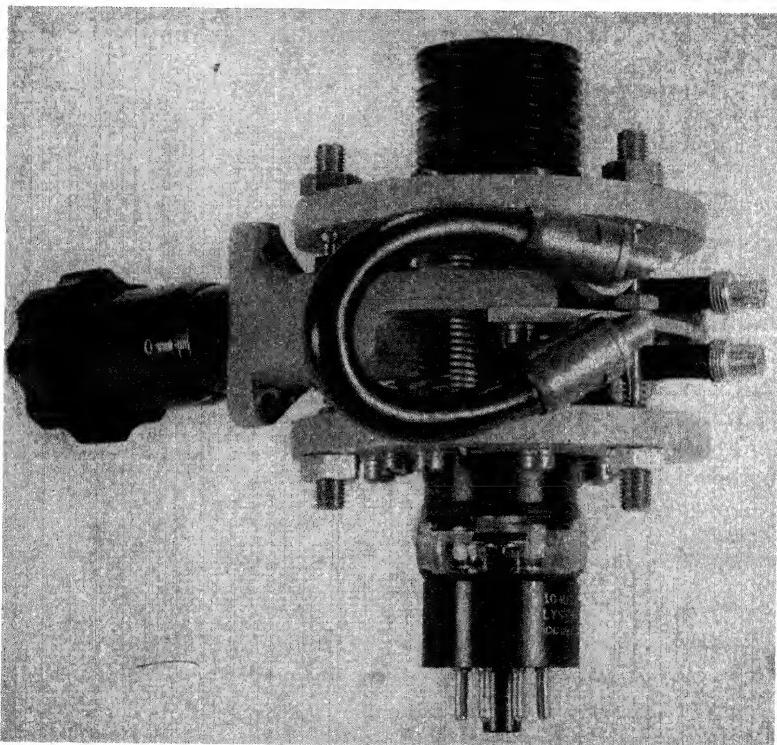
FIG. 6.—Double-resonator klystron with tuning mount. (*Courtesy of the Sperry Gyroscope Company.*)

In the operation of the klystron, the electrons are accelerated by the d-c field resulting from the potential V_0 and consequently enter the buncher-grid region with a high initial velocity. If the klystron is operating as an amplifier, the buncher resonator is excited at its resonant frequency by the incoming signal which is to be amplified. This produces an alternating field between the buncher grids. As electrons pass through this field they are either accelerated or decelerated, depending upon the phase of the buncher voltage during their transit. Those electrons which pass through the buncher grids during the accelerating phase are speeded

up and emerge with a velocity higher than the entering velocity. Other electrons, traveling through the buncher grids during the retarding phase, are slowed down and emerge from the buncher grids with reduced velocity. This variation of velocity of the electrons in an electron stream is known as *velocity modulation.*

In the field-free drift space between buncher and catcher grids, the high-velocity electrons overtake the low-velocity electrons which left the buncher grids at an earlier phase. This results in a bunching of the elec-
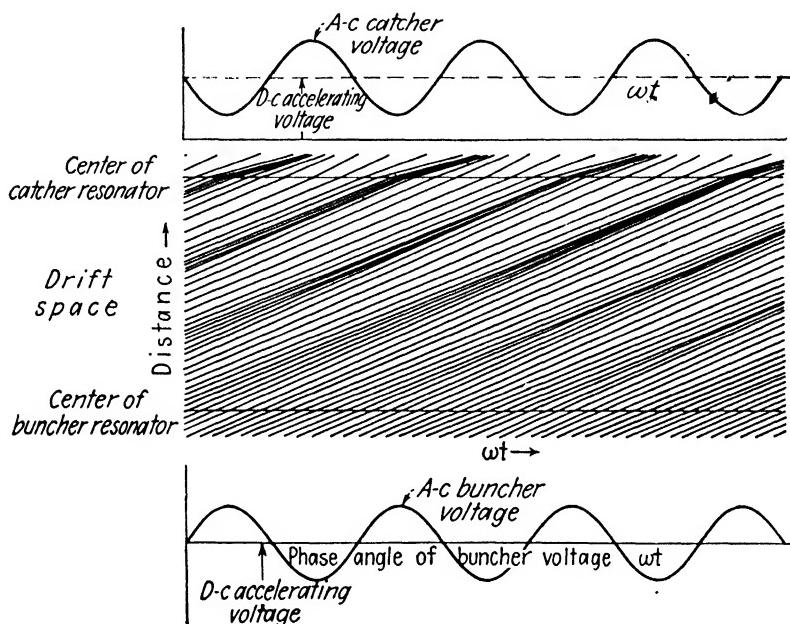
FIG. 7.—Applegate diagram representing electron bunching in the klystron.

trons as they drift toward the catcher resonator. For optimum performance, maximum bunching should occur approximately midway between the catcher grids. The electron bunches pass through the alternating field between the catcher grids during its retarding phase; hence energy is transferred from the electrons to the field of the catcher resonator. The electrons emerge from the catcher grids with reduced velocity and finally terminate at the collector. If the klystron is operating as an oscillator, energy is fed back from the catcher to the buncher through a short coaxial line in order to produce sustained oscillation.

The bunching process is illustrated by the Applegate diagram of Fig. 7. This shows the electron displacement as a function of time for electrons leaving the buncher at different phases of buncher voltage. Each line represents the displacement-time relationship for a single electron. The

slope of the line is proportional to electron velocity; hence the higher velocity electrons are represented by lines having steeper slopes. The electron bunches center around the electron which passes through the buncher grids when the buncher voltage is zero and changing from deceleration to acceleration. This particular electron will be designated the center-of-the-bunch electron.
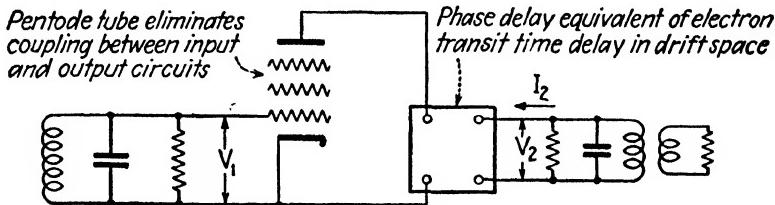
FIG. 8.—Equivalent circuit of a klystron.

The klystron may be represented by the equivalent circuit of Fig. 8. The multigrid tube emphasizes the complete isolation of input (buncher) and output (catcher) circuits. The delay network represents the buncher-to-catcher transit-time delay. The two tuned circuits represent buncher and catcher resonators. The load is shown coupled to the output or catcher resonator. The current I_2 is the induced current flowing in the catcher resonator. While the equivalent circuit is helpful in visualizing the over-all functioning of the system. the analysis must necessarily proceed along more fundamental lines.

(a)   (b)   (c)
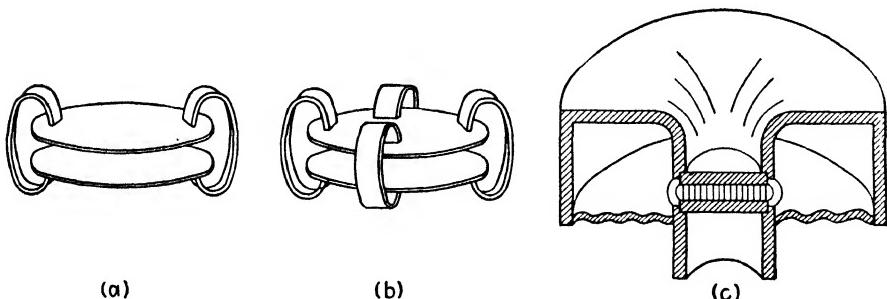
FIG. 9.—Development of the resonator.

6.05. The Klystron Resonator.—The phenomenon of electromagnetic oscillation in resonators may be illustrated by the development of the resonator in Fig. 9. Figure 9a shows a parallel resonant circuit consisting of a parallel-plate condenser and two inductive loops. Adding more turns in parallel, as shown in **Fig.** 9b, decreases the inductance and increases the resonant frequency. Carrying this to the limit, we have the

totally enclosed toroidal resonator of Fig. 9c, with the electromagnetic field confined entirely to the inside of the resonator.

The effective capacitance of the resonator of Fig. 9c is approximately equal to the capacitance of the parallel grids, while the inductance is roughly proportional to the volume of the resonator. Hence, the resonant frequency may be increased either by increasing the separation distance between grids (decreasing the capacitance) or decreasing the volume of the resonator (decreasing the inductance). Tuning is accomplished in the klystron oscillator by making one wall of the resonator a thin corrugated diaphragm. Pressure on this diaphragm varies the spacing between resonator grids, thereby changing the effective capacitance of the resonator The tuning mechanism is shown attached to the klystron in Fig. 6.

6.06. Electron Transit-time Relationships in the Klystron.[1-7]—The analysis of the klystron is based upon the following simplifying assumptions:

1. The electron transit angles through the buncher and catcher grids are assumed to be negligible.[8]

2. Space-charge effects are ignored.

3. The alternating voltage between buncher grids is assumed to be small in comparison with the d-c accelerating voltage.

4. The electron beam is assumed to have uniform density in the cross section of the beam, and all of the electrons which leave the cathode are assumed to pass through the catcher grids.

These assumptions lead to idealistic values of power output and efficiency which are considerably greater than those realized in practice. We shall derive the theoretical relationships and then discuss the conditions prevailing in practice.

In the klystron oscillator of Fig. 10, the electrons are accelerated to an initial velocity v_0 by the d-c field before they enter the buncher grids.

[1] VARIAN, R. H., and S. H. VARIAN, High Frequency Oscillator and Amplifier, *J. Applied Phys.*, vol. 10, pp. 321–327; May, 1939.

[2] WEBSTER, D. L., Cathode-Ray Bunching, *J. Applied Phys.*, vol. 10, pp. 501–513; July, 1939.

[3] WEBSTER, D. L., The Theory of Klystron Oscillations, *J. Applied Phys.*, vol. 10, pp. 864–872; December, 1939.

[4] HARRISON, A. E., "Klystron Technical Manual," Sperry Gyroscope Co., Brooklyn, New York, 1944.

[5] HARRISON, A. E., Klystron Oscillators, *Electronics*, vol. 17, pp. 100–107; November, 1944.

[6] CONDON, E. U., Electronic Generation of Electromagnetic Oscillations, *J. Applied Phys.*, vol. 11, pp. 502–507; July, 1940.

[7] CONDON, E. U., Microwave Generators Using Velocity-Modulated Electron Beams, *Proc. National Electronics Conference*, vol. 1, pp. 500–513; October, 1944.

[8] The transit angle of electrons through the buncher or catcher may be quite large for low accelerating voltages, consequently the assumption of negligible transit angle may be a poor approximation.

Assume that the electrons leave the cathode with zero velocity. The kinetic energy and the velocity of the electrons entering the buncher, as given by Eqs. (2.06-5 and 7), are

$$\tfrac{1}{2}mv_0^2 = eV_0 \tag{1}$$

$$v_0 = \sqrt{\frac{2eV_0}{m}} \tag{2}$$

Oscillation of the buncher resonator produces an alternating potential difference $V_1 \sin \omega t$ between resonator grids. It is assumed that the two buncher grids in Fig. 10 are at potentials V_0 and $V_0 + V_1 \sin \omega t$, both
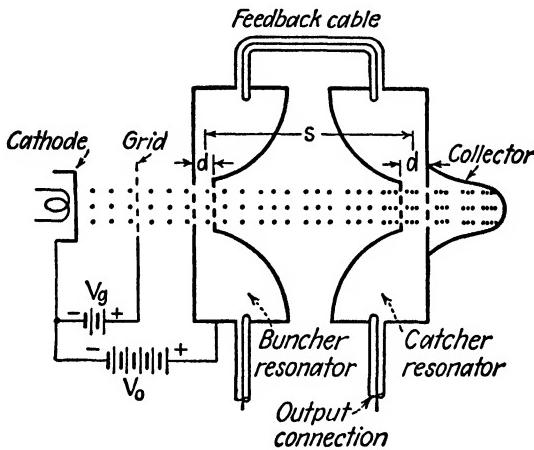
FIG. 10.—Klystron oscillator.

with respect to the cathode. Consider an electron passing through the buncher at time t_1. Assume that the buncher voltage remains constant during the passage of any one electron through the buncher grids. The kinetic energy $\tfrac{1}{2}mv_1^2$ and velocity v_1 of the electron as it emerges from the buncher grids are

$$\tfrac{1}{2}mv_1^2 = e(V_0 + V_1 \sin \omega t_1)$$

$$v_1 = \sqrt{\frac{2e}{m} V_0 \left(1 + \frac{V_1}{V_0} \sin \omega t_1\right)} = v_0 \sqrt{1 + \frac{V_1}{V_0} \sin \omega t_1} \tag{3}$$

Equation (3) is the equation of velocity modulation. An electron which passes through the buncher grids when $\omega t_1 = 0$ emerges with unchanged velocity. The maximum and minimum velocities at the buncher-grid exit are $v_0 \sqrt{1 + (V_1/V_0)}$ and $v_0 \sqrt{1 - (V_1/V_0)}$, these representing the velocities of electrons which pass through the buncher grids during the maxi-

num accelerating and maximum retarding phases of the buncher voltage; *i.e.*, when $\omega t_1 = \pi/2$ and $\omega t_1 = -\pi/2$, respectively.

The time T required for the electron to travel the distance s in the field-free space between buncher and catcher grids is

$$T = \frac{s}{v_1} = \frac{s}{v_0 \sqrt{1 + \dfrac{V_1}{V_0} \sin \omega t_1}} \tag{4}$$

Using the buncher alternating voltage as the time reference, the electron arrives at the catcher grids at time $t_2 = t_1 + T$. The corresponding phase angle (with the buncher voltage as reference) is obtained by multiplying both sides of the equation for t_2 by ω. Substitution of Eq. (4) yields

$$\omega t_2 = \omega(t_1 + T)$$

$$= \omega t_1 + \frac{\omega s}{v_0}\left(1 + \frac{V_1}{V_0} \sin \omega t_1\right)^{-\frac{1}{2}} \tag{5}$$

The quantity s/v_0 is the buncher-to-catcher transit time for the electron passing through the buncher when $\omega t_1 = 0$, *i.e.*, the center-of-the-bunch electron. The corresponding transit angle is

$$\alpha = \frac{\omega s}{v_0} \tag{6}$$

In subsequent derivations, mathematical difficulties are encountered if we use Eq. (5) in its present form. To avoid complication, we assume that $V_1/V_0 \ll 1$ and expand the bracketed term into a binomial series. Taking the first two terms of the series and assuming that the remaining terms are negligible, we obtain $[1 + (V_1/V_0) \sin \omega t_1]^{-\frac{1}{2}} \approx 1 - (V_1/2V_0) \sin \omega t_1$. Substituting this, together with Eq. (6), in Eq. (5), we obtain

$$\omega t_2 = \omega t_1 + \alpha\left(1 - \frac{V_1}{2V_0} \sin \omega t_1\right) \tag{7}$$

The interpretation of Eq. (7) is as follows: An electron leaving the buncher at phase angle ωt_1 (with respect to the buncher voltage) arrives at the catcher at phase angle ωt_2 (also measured with respect to the buncher voltage). The arrival phase angle ωt_2 and departure phase angle ωt_1 are related by Eq. (7).

6.07. Power Output and Efficiency of the Klystron.—Now consider the energy and power transfer at the catcher resonator. The analysis is simplified by assuming that the buncher and catcher voltages are in time phase and therefore the catcher voltage will be represented by $V_2 \sin \omega t$. If the catcher voltage is substantially constant during an electron's transit

through the catcher grids, the electron gives up an amount of energy $w = -eV_2 \sin \omega t_2$. The negative sign signifies energy transfer from the electron to the field. Using this convention, energy output and power output are positive quantities. Substituting the value of ωt_2 from Eq. (6.06-7) into the energy output equation, we obtain

$$w = -eV_2 \sin \omega t_2$$

$$= -eV_2 \sin \left[\omega t_1 + \alpha \left(1 - \frac{V_1}{2V_0} \sin \omega t_1 \right) \right] \tag{1}$$

Averaging Eq. (1) for all electrons entering the buncher grids during a complete cycle, *i.e.*, between $\omega t_1 = 0$ and $\omega t_1 = 2\pi$, we obtain the average energy per electron transferred to the catcher resonator,

$$w_{\mathrm{av}} = \frac{1}{2\pi} \int_{\omega t_1=0}^{\omega t_1=2\pi} w \, d(\omega t_1)$$

$$= -\frac{eV_2}{2\pi} \int_0^{2\pi} \sin \left[\omega t_1 + \alpha \left(1 - \frac{V_1}{2V_0} \sin \omega t_1 \right) \right] d(\omega t_1) \tag{2}$$

This integral yields [1]

$$w_{\mathrm{av}} = -eV_2 J_1(x) \sin \alpha \tag{3}$$

where

$$x = \frac{\alpha V_1}{2V_0} \tag{4}$$

The quantity x features prominently in klystron analysis and is known as the *bunching parameter*. The quantity $J_1(x)$ is a Bessel function of the first kind and first order. The Bessel function may be expressed as an

---

[1] Substituting $x = \alpha V_1/2V_0$ and expanding, we obtain for the integrand of Eq. (2),

$$\sin (\omega t_1 + \alpha - x \sin \omega t_1) = \sin (\omega t_1 + \alpha) \cos (x \sin \omega t_1) - \cos (\omega t_1 + \alpha) \sin (x \sin \omega t_1)$$

In standard treatments of Bessel's functions it is shown that

$$\cos (x \sin \omega t_1) = J_0(x) + 2J_2(x) \cos 2\omega t_1 + 2J_4(x) \cos 4\omega t_1 + \cdots$$

$$\sin (x \sin \omega t_1) = 2J_1(x) \sin \omega t_1 + 2J_3(x) \sin 3\omega t_1 + \cdots$$

When the first series is multiplied by $\sin (\omega t_1 + \alpha)$ and integrated between the limits $\omega t_1 = 0$ and $\omega t_1 = 2\pi$, the integral has zero value since each term is of the form $\int_0^{2\pi} J_n(x) \sin (\omega t_1 + \alpha) \cos n\omega t_1 \, d(\omega t_1)$ where n is an even integer. When the second series is multiplied by $-\cos (\omega t_1 + \alpha)$ and integrated, each of the terms is of the form $-\int_0^{2\pi} J_n(x) \cos (\omega t_1 + \alpha) \sin n\omega t_1 \, d(\omega t_1)$. All of the terms except that corresponding to $n = 1$ are zero. The remaining term is

$$w_{\mathrm{av}} = \frac{eV_2}{2\pi} \int_0^{2\pi} 2J_1(x) \sin \omega t_1 \cos (\omega t_1 + \alpha) \, d(\omega t_1) = -eV_2 J_1(x) \sin \alpha$$

infinite series in terms of its argument x as given in Eq. (15.05-12). A plot of $J_1(x)$ as a function of x is shown in Fig. 11.

Let us now consider the power relationships. Assume that N electrons are emitted from the cathode per second. The direct current emitted from the cathode is then $I_0 = Ne$. (Positive current is assumed to flow in the direction opposite to that of electron flow.) Now assume that all of the electrons emitted from the cathode flow through the catcher resonator; consequently there are N electrons flowing through the catcher per second. Equation (3) expresses the average energy transfer per electron.
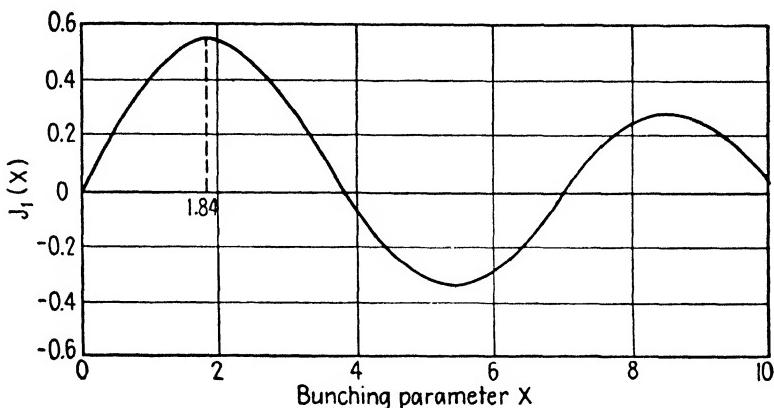
FIG. 11.—Plot of $J_1(x)$ against x.

To obtain the energy transfer per second or power transfer, we multiply Eq. (3) by N. Substituting $I_0 = Ne$ we obtain for the power output P_{ac} at the catcher resonator

$$P_{ac} = -I_0 V_2 J_1(x) \sin \alpha \qquad (5)$$

The power supplied by the d-c potential source is $P_0 = I_0 V_0$ and the conversion efficiency is

$$\eta = \frac{P_{ac}}{P_0} = -\frac{V_2}{V_0} J_1(x) \sin \alpha \qquad (6)$$

In Eq. (6) it is assumed that the power required for bunching the electrons is negligible. Since the alternating field of the buncher accelerates some electrons and retards others, the time-average power required for the bunching operation would be zero if the electron transit angle were negligible. In actual practice, however, the transit angle is not negligible and some power is required to perform the bunching operation.

6.08. Requirements for Maximum Output and Maximum Efficiency.—
The Bessel function $J_1(x)$ appears in both the power output and the efficiency equation. Figure 11 shows that $J_1(x)$ has a maximum value of 0.58

when $x = 1.84$. Referring to Eq. (6.07-6), the conditions required for maximum efficiency are found to be

1. A maximum value of the ratio V_2/V_0.
2. A bunching parameter value of $x = 1.84$, yielding $J_1(x) = 0.58$.
3. $\sin \alpha = -1$, or $\alpha = 2\pi n - (\pi/2)$ where n is any positive integer greater than zero.

In general, the value of V_2/V_0 is less than unity. If V_2 were greater than V_0 some of the electrons would reverse their direction of travel at the catcher resonator and absorb power from the alternating field. If we assume a value $V_2/V_0 = 1$ and $\sin \alpha = -1$, the maximum theoretical efficiency becomes 58 per cent. In addition to the requirements for maximum efficiency, maximum power output requires that the d-c beam current I_0 and the catcher resonator potential V_2 have maximum possible values.

It will be shown later that, for the assumed conditions, criterion 3 above is always satisfied if sustained oscillations exist. Equations (6.07-5 and 6) therefore become

$$P_{ac} = I_0 V_2 J_1(x) \tag{1}$$

$$\eta = \frac{V_2}{V_0} J_1(x) \tag{2}$$

The criterion $\alpha = 2\pi n - (\pi/2)$ determines the value of the d-c accelerating voltage required for maximum power output. To obtain this relationship, we return to Eq. (6.06-6) and set $\alpha = 2\pi n - (\pi/2)$ and $v_0 = \sqrt{2eV_0/m}$, yielding

$$V_0 = \left(\frac{\omega s}{2\pi n - (\pi/2)}\right)^2 \frac{m}{2e} = 0.284 \times 10^{-11} \left(\frac{\omega s}{2\pi n - (\pi/2)}\right)^2 \tag{3}$$

Equation (3) reveals that there are a multiplicity of values of d-c accelerating voltage which will yield maximum power output (one for each integral value of n). The highest voltage corresponds to the lowest integer; *i.e.*, $n = 1$.

The criterion $x = 1.84$ required for maximum power output and maximum efficiency determines the optimum ratio of buncher voltage to d-c accelerating voltage V_1/V_0. To show this we substitute $\alpha = 2\pi n - (\pi/2)$ and $x = 1.84$ in Eq. (6.07-4) obtaining

$$\frac{V_1}{V_0} = \frac{3.68}{2\pi n - (\pi/2)} \tag{4}$$

For each integral value of n there is an optimum value of V_1/V_0 which yields maximum power output. When $n = 1$ we obtain $V_1/V_0 = 0.78$. As n increases, the d-c accelerating voltage V_0 and the ratio V_1/V_0 both decrease. The reason for this becomes apparent when we recall that the

electron velocity varies as the square root of V_0. Consequently, if V_0 is small, the electrons travel more slowly and have more time in which to bunch; therefore less bunching voltage is required.

In the normal operation of the klystron oscillator, electron bunching occurs only at the catcher resonator. However, it is possible to have a condition of oscillation in which the electrons alternately bunch and debunch several times in the drift space between buncher and catcher resonators. This is indicated by the successive maxima of $J_1(x)$ for increasing values of x in Fig. 11. The value $x = 1.84$ corresponds to single bunching and yields the highest efficiency.
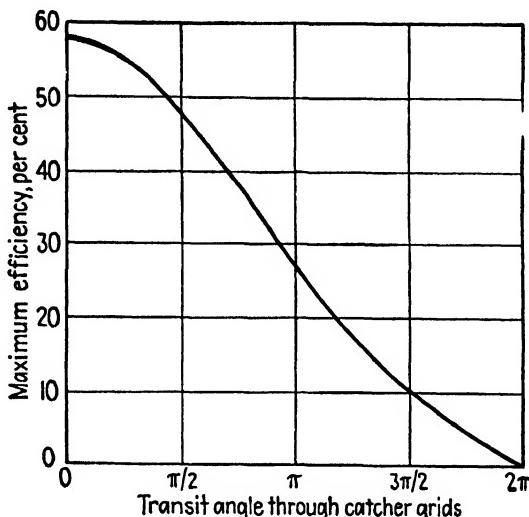
FIG. 12.—Maximum efficiency of the klystron as a function of catcher transit angle.

Maximum power output requires a maximum value of the catcher resonator voltage V_2. The catcher voltage increases with the d-c accelerating voltage (this is not apparent from the above equations), and hence the highest permissible accelerating voltage yields maximum power output if all other conditions are satisfied.

The voltage V_2 is also dependent upon the effective impedance of the catcher resonator. The resonator is analogous to a parallel *L-C* circuit. At its resonant frequency, the impedance of the resonator is a pure resistance with a value proportional to the Q of the resonator and to its effective L/C ratio. In order to obtain maximum voltage V_2, the resonator impedance should be as large as possible. However, there are other important considerations involved which necessitate design compromises.

Klystron resonators must be designed with a view toward minimizing electron transit time through the resonator grids. Excessive electron

transit time through the buncher grids results in inadequate electron bunching and power is consumed in the bunching process. If the transit time through the catcher resonator is excessive, the power output and efficiency are reduced. Figure 12 shows how the maximum theoretical efficiency of the klystron decreases as the catcher-resonator transit angle increases.[1] Electron transit angle is minimized by spacing the grids close together. However, this increases the capacitance and lowers the effective impedance of the resonator. To avoid excessive transit time and still obtain high impedance, a compromise in the spacing between resonator grids is necessary.
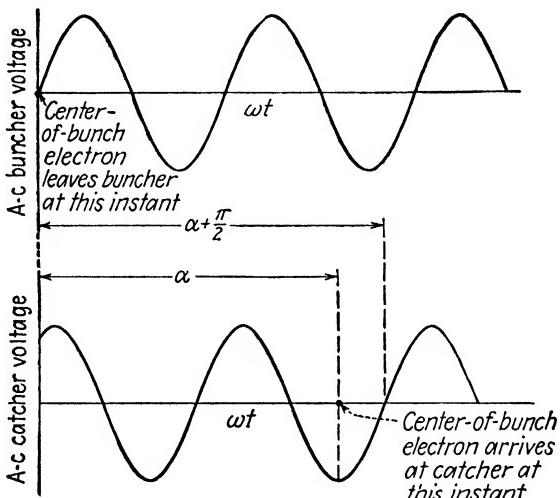
FIG. 13.—Phase relationships in the double-resonator klystron.

6.09. Phase Relationships in the Klystron Oscillator.—The foregoing analysis was simplified by the assumption that the buncher and catcher resonators oscillate in time phase. If this were a necessary restriction, it would result in extremely critical operation and the slightest deviation of the voltages from the values given in Eqs. (6.08-3 and 4) would stop oscillation. Actually, however, there is a certain amount of flexibility permissible in the phase relationships between buncher and catcher voltages as a result of phase shift existing in the resonators and feedback line.

In Eq. (5.03-3) a criterion of oscillation was stated in the form $\beta K = 1$ where K is the voltage gain; *i.e.*, the ratio of output to input voltage, and β is the ratio of feedback voltage to output voltage. A more general expression for this criterion is $\beta K = 1/2\pi n$ radians, where n is any integer in-

[1] BLACK, L. J., and P. L. MORTON, Current and Power in Velocity-Modulation Tubes, *Proc. I.R.E.*, vol. 32, pp. 477–482; August, 1944.

cluding zero. This requires that the sum of the phase angles of K and β, or the total phase shift around the closed circuit, be equal to $2\pi n$ radians.

Let us now apply this criterion to determine the phase relationships in the klystron oscillator. The diagram of Fig. 13 is drawn for the general case in which buncher and catcher voltages are not in time phase. The center-of-the-bunch electron leaves the buncher when its alternating voltage is zero. This electron arrives at the catcher resonator when its alternating voltage has its maximum negative value. The transit angle α, given by Eq. (6.06-6), is the phase angle between the zero of the buncher
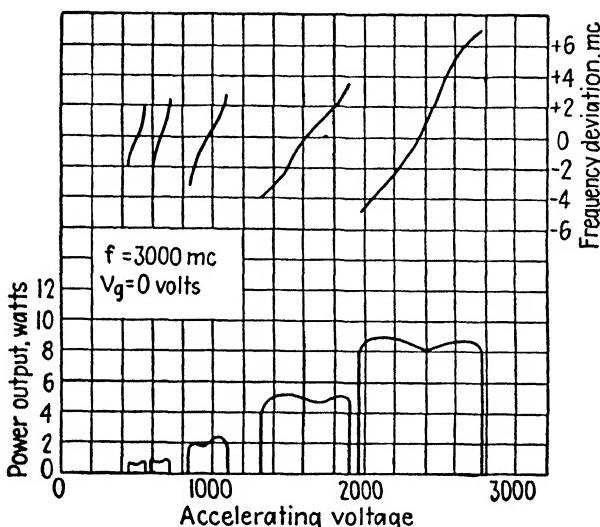
FIG. 14.—Experimental curves showing power output and frequency deviation as a function of d-c accelerating voltage for a double-resonator klystron.

voltage and the negative maximum of the catcher voltage. Consequently, the phase angle between the zeros of buncher and catcher voltage is $\alpha + (\pi/2)$ radians. In adding up the phase shifts around a closed circuit in a klystron, we must include the phase shift in the resonators and in the feedback line. If a resonator is operating at its resonant frequency, the voltage and current are in time phase and the phase shift in the resonator is zero. However, the frequency of oscillation of a klystron may deviate slightly from the resonant frequency of the resonator. In this case, the voltage and current are not in time phase and there is a phase shift in the resonator. If we let θ be the total phase shift in the resonators and feedback cable, the criterion of oscillation requires that

$$\theta + \alpha + \frac{\pi}{2} = 2\pi n \qquad \text{radians} \qquad \cdot(1)$$

In the derivations of Secs. 6.06 and 6.07, it was assumed that the two resonators oscillate in time phase. This corresponds to setting $\theta = 0$ in Eq. (1). We found that maximum power output was obtained when $\alpha = 2\pi n - (\pi/2)$. We now find that if the two resonators oscillate in time phase, the condition that $\alpha = 2\pi n - (\pi/2)$ is not only a requirement for maximum power output but is also a criterion which must exist if sustained oscillations are to be obtained. In general, however, the resonators do not have to oscillate in time phase.

In the operation of the klystron oscillator, it is found that a small change in d-c accelerating voltage will cause a change in frequency. This can be
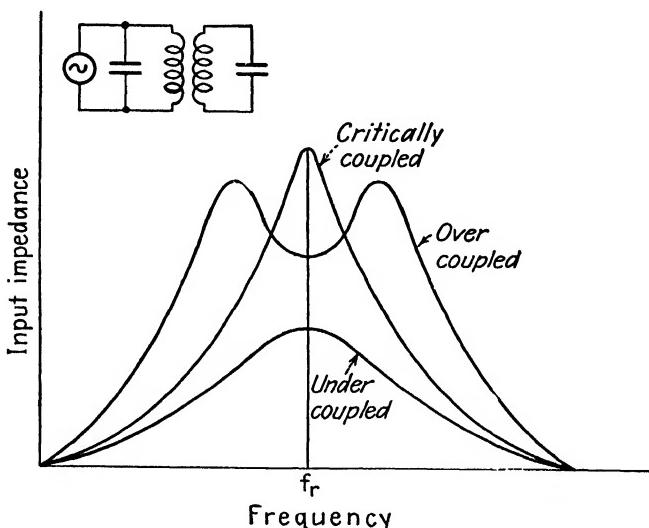
FIG. 15.—Input impedance of two coupled circuits having the same resonant frequency.

explained by the fact that the change in d-c voltage causes a variation in the transit angle α. The frequency of oscillation then shifts in such a way as to yield a new value of θ which will satisfy Eq. (1).

It is a well-known principle in circuit theory that if two tuned circuits having the same resonant frequency are coupled together the input impedance looking into either circuit will have a variation with frequency such as that shown in Fig. 15. A double-peaked resonance curve occurs when the tuned circuits are overcoupled, while the single-peaked curve corresponds to critical coupling or undercoupling. The same phenomenon occurs when two resonators are coupled together.

It is possible to obtain oscillation over a somewhat wider range of d-c accelerating voltages if the resonators are overcoupled. It is interesting to note that the critically-coupled klystron has practically a straight-line variation in frequency with accelerating voltage, thereby offering attractive

possibilities for frequency modulation. By tuning the resonators to slightly different frequencies, the power output may be held more nearly constant. The curves of Fig. 14 were obtained with the two resonators tuned to slightly different frequencies.

6.10. Current and Space-charge Density in the Klystron.—An interesting interpretation of the bunching process may be obtained by plotting curves of electron departure time at the buncher against arrival time at
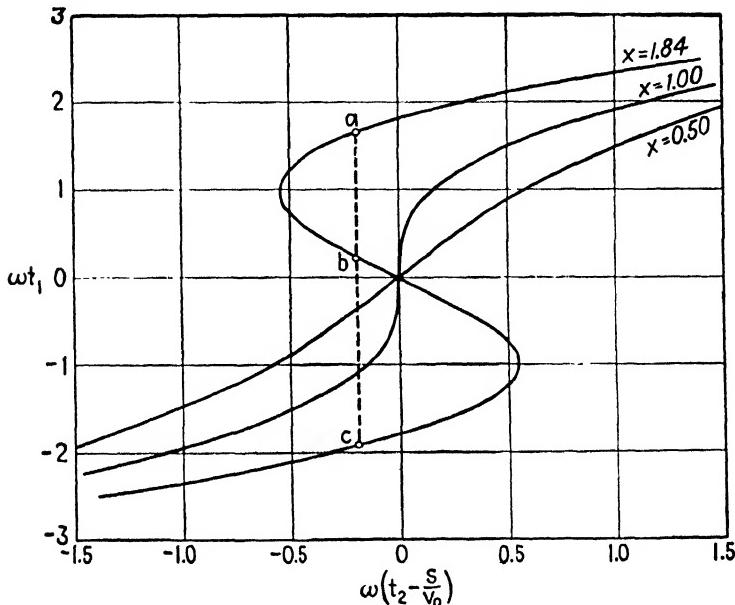
Fig. 16.—Plot of electron departure phase angle against arrival phase angle.

the catcher. We start by substituting Eqs. (6.06-6) and (6.07-4) into (6.06-7) and rewriting this in the form

$$\omega\left(t_2 - \frac{s}{v_0}\right) = \omega t_1 - x \sin \omega t_1 \tag{1}$$

In this equation ωt_1 is the phase angle at which an electron departs from the buncher and ωt_2 is the phase angle at which this electron arrives at the catcher, both measured with respect to the buncher voltage.

Curves of ωt_1 as a function of $\omega[t_2 - (s/v_0)]$ for various values of the bunching parameter x are shown in Fig. 16. The quantity $\omega[t_2 - (s/v_0)]$ is the arrival phase angle at the catcher minus a constant amount $\alpha = \omega s/v_0$ which represents the buncher-to-catcher transit angle of the center-of-the-bunch electron. This plot shows that the electrons arriving at the catcher at any instant of time may have left the buncher at different instants of time. Thus, referring to the curve $x = 1.84$ in Fig. 16, we find

that electrons leaving the buncher at times corresponding to a, b, and c all arrive simultaneously at the catcher resonator.

The space-charge densities at the buncher and catcher resonators will be represented by $q_{\tau 1}$ and $q_{\tau 2}$, and the corresponding velocities are v_1 and v_2, respectively. The amount of charge flowing through unit area at the buncher exit during the time interval dt_1 is $q_{\tau 1} v_1 \, dt_1$. This same charge flows through unit area at the catcher during a different time interval dt_2, and may be expressed as $q_{\tau 2} v_2 \, dt_2$; hence we have

$$q_{\tau 1} v_1 \, dt_1 = q_{\tau 2} v_2 \, dt_2 \qquad (2)$$

Equation (2) is not valid if we have a condition in which some electrons overtake other electrons in the drift space. If this occurs, the charges which arrive at the catcher during the time interval dt_2 may have left the buncher at several different intervals of dt_1. Hence, there may be several different values of $q_{\tau 1} v_1 \, dt_1$ which contribute to the charge $q_{\tau 2} v_2 \, dt_2$. To take care of this situation, we express $q_{\tau 2} v_2 \, dt_2$ as the summation of all of the values of $q_{\tau 1} v_1 \, dt_1$ which contribute to it, thus

$$q_{\tau 2} v_2 \, dt_2 = \Sigma q_{\tau 1} v_1 \, dt_1 \qquad (3)$$

The space charge density $q_{\tau 1}$ at the buncher is approximately equal to the space charge density $q_{\tau 0}$ just before the buncher; therefore we let $q_{\tau 1} = q_{\tau 0}$. Also, if the velocity variation is small, we have $v_2 \approx v_1 \approx v_0$. Making these substitutions in Eq. (3) and rearranging, we obtain

$$q_{\tau 2} v_0 \, dt_2 = q_{\tau 0} v_0 \Sigma \, dt_1$$

$$\frac{q_{\tau 2}}{q_{\tau 0}} = \Sigma \frac{dt_1}{dt_2} = \Sigma \frac{d(\omega t_1)}{d(\omega t_2)} \qquad (4)$$

The quantity $d(\omega t_1)/d(\omega t_2)$ is the slope of the curve of Fig. 16. Thus, to obtain the space-charge density ratio $q_{\tau 2}/q_{\tau 0}$ for any given value of $\omega[t_2 - (s/v_0)]$ we need merely add the slopes of the curve for the given value of $\omega[t_2 - (s/v_0)]$. A negative slope indicates that electrons which arrive at the catcher in one sequence left the buncher in the reverse sequence. In adding the slopes, we consider only absolute values and discard the sign. The curves of Fig. 17 were obtained in this manner.

The space-charge density curves plotted in Fig. 17 are not what we might have anticipated. The value of $x = 1.84$, which we previously found to yield maximum power output, is not the critically bunched condition. The double-peak in this curve indicates overbunching with two distinct groups of electrons flowing through the catcher resonator a short time interval apart. The critically bunched condition, *i.e.*, the condition of maximum bunching, corresponds to $x = 1.0$. Since the power output and efficiency vary directly as $J_1(x)$, Fig. 11 shows that the maximum theo-

retical power output and efficiency for the critically bunched condition is 24 per cent lower than for the overbunched value of $x = 1.84$.

The convection-current density at any point in the electron stream is the product of charge density times velocity. The convection-current density at the buncher and catcher resonators is $J_1 = q_{r1}v_1 \approx q_{r0}v_0$ and $J_2 = q_{r2}v_2 \approx q_{r2}v_0$. The convection-current density at the buncher is practically equal to the current density J_0 just before the buncher; hence we
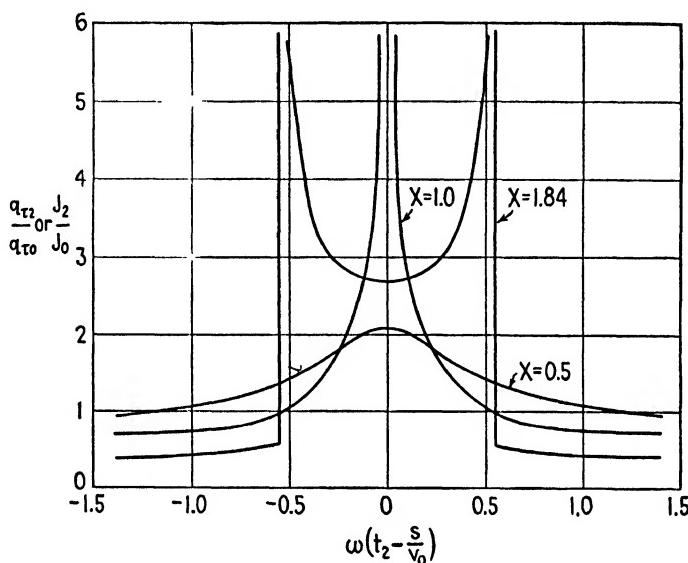
FIG. 17.—Plot of q_{r2}/q_{r0} and J_2/J_0 as a function of arrival phase angle.

may substitute $J_1 = J_0$. Making this substitution and dividing the above equations, we obtain

$$\frac{J_2}{J_0} = \frac{q_{r2}}{q_{r0}} \tag{5}$$

We find, therefore, that the curves of Fig. 17 may be used to represent either the charge-density ratio q_{r2}/q_{r0} or the convection-current density ratio J_2/J_0.

The beam current has a high harmonic content. The current induced in the catcher resonator has a similar waveform. This current can be analyzed by Fourier series to obtain the d-c component, fundamental a-c component, and higher harmonic components. The catcher resonator may be tuned to any harmonic of the buncher frequency. The a-c components of current have amplitudes proportional to $J_m(mx)$, where m is the order

for the mth harmonic.   If V_m and I_m are the peak values of the mth harmonic voltage and induced current at the catcher resonator, the time-average power output is $P_{ac} = V_m I_m / 2$.   Equating this to the power output obtained previously, and solving for I_m, we obtain

$$I_m^- = 2I_0 J_m(mx) \tag{6}$$

The effective impedance of the resonator at the mth harmonic of the buncher voltage is

$$Z_m = \frac{V_m}{I_m} \tag{7}$$

If the catcher resonator is operating at its resonant frequency, the resonator impedance is a pure resistance.

When the klystron is used as an oscillator, the buncher and catcher resonators are tuned to the same frequency and we have $m = 1$ in the above equations.   The klystron may be used as a frequency multiplier, in which case the catcher is tuned to a harmonic of the buncher voltage.

The resonant impedance of a resonator is determined by the geometry of the resonator, the loading, and the electronic effects due to the beam current.   In general, the impedance is increased by increasing the effective L/C ratio, which means increasing the volume and decreasing the capacitance between grids.

6.11. Operation of the Klystron.—In the tuning of the klystron it is necessary to make simultaneous adjustment of a number of variables. These include the grid voltage, the d-c accelerating voltage, and the tuning of the two resonators.   The tuning procedure may be simplified by connecting a 60-cycle alternating voltage of from 50 to 100 volts in series with the d-c accelerating voltage.   The accelerating voltage may then be adjusted to the approximate value required for oscillation and left at this value while the remaining adjustments are made.   The presence of oscillation may be observed by means of a crystal detector and microammeter connected to the klystron output.   The 60-cycle a-c voltage is removed during the final adjustment..

Power outputs of the order of a fraction of a watt to several hundred watts are obtainable with klystrons at wavelengths of the order of 10 centimeters or less.   Actual efficiencies run far below the ideal efficiency of 58 per cent.   This may be attributed to space-charge effects causing debunching of the electrons, collisions of electrons with the grids, secondary emission at the grid, power consumed in bunching the electrons, transit-angle delay of electrons in their passage through buncher and catcher grids, and losses in the resonators.

The frequency stability of the klystron oscillator is dependent upon the temperature of the resonator as well as the stability of the power-supply

voltages.   Since the resonator walls are metallic, their thermal coefficient of expansion is quite large; hence the frequency drift with temperature may be appreciable.   By the choice of suitable materials, it is possible to make the thermal expansion of the tuning mechanism partially compensate for the expansion of the resonator walls.   Carefully regulated power supplies are required for high frequency stability.

The double-resonator klystron may be used as an amplifier [1] by coupling the input signal into the buncher resonator and obtaining the output from the catcher resonator.   Regenerative amplification may be obtained by

Output terminal

Catcher grids

Cascade buncher grids

Buncher grids

Input terminal
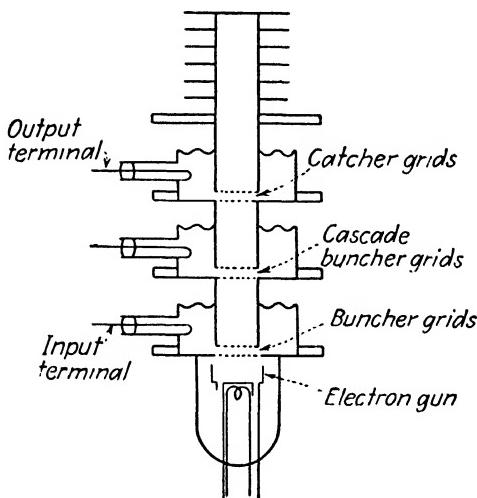
Electron gun

FIG. 18.—Cascade klystron.

loosely coupling the buncher and catcher resonators.   The adjustments of the klystron as an amplifier are less critical than as an oscillator.   It is not necessary for the electrons to be entirely bunched when the klystron is operating as an amplifier.

Higher amplification may be obtained by the use of a cascade system of resonators as shown in Fig. 18.   When the cascade klystron is operated as an amplifier, the input signal is fed into the buncher resonator.   Partial bunching occurs at the second resonator and the a-c field (produced by the electron bunches passing through this resonator) velocity modulates the electron stream in such a manner as to increase the electron bunching. The cumulative bunching of several resonators makes it possible to approach critical bunching at the final resonator even though the input signal is relatively weak.   Power gains of 2 to 20 per stage are attainable.   A

[1] HAXBY, R. O., The Principles of Klystron Amplifiers, *Proc. National Electronics Conference*, vol. I, pp. 229–240; October, 1944.

cascade klystron may consist of two stages as an oscillator and the third stage as an amplitude-modulated amplifier. The general problem of modulation of microwave systems is considered in Chap. 11.

6.12. The Reflex Klystron.—The reflex klystron, illustrated in Fig. 19, employs a single resonator which combines the functions of buncher and catcher resonators. A negative-potential reflector electrode, placed a short distance beyond the resonator grids, sets up a retarding d-c field which causes the electrons to reverse their direction of travel and proceed back through the resonator grids. If oscillating conditions prevail, the electrons are velocity modulated on their first excursion through the resonator grids. The electrons then enter the d-c retarding field in the reflecting space where they are turned back and proceed to pass through the resonator grids in the opposite direction. Electron bunching occurs during the transit through the reflecting space. On their return journey, the electron bunches pass through the alternating field between the resonator grids during its retarding phase and thus energy is transferred from the electrons to the resonator alternating field.
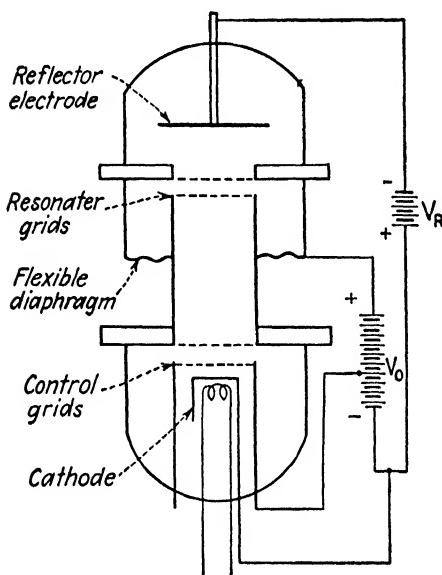
FIG. 19.—Reflex klystron.

Bunching action in the reflex klystron differs from that of the double-resonator klystron in several respects. In the double-resonator klystron, all electrons travel the same distance between buncher and catcher and the bunch centers around the electron which passes through the buncher grids when its field is changing from deceleration to acceleration. In the reflex klystron, the higher velocity electrons travel farther in the reflecting space and take longer to return. For this reason, bunching centers around the electron which passes through the grids on its first excursion when the resonator alternating potential is zero, changing from acceleration to deceleration. Furthermore, in the reflex klystron, the returning electrons flow through the resonator grids in a direction opposite to the direction of travel through the catcher grids of the double-resonator klystron. Consequently there is a 180-degree phase difference between the resonator voltage of the reflex klystron and the catcher voltage of the double-resonator klystron at the instant when the bunched electrons pass through the resonator grids.

The type of bunching encountered in the double-resonator klystron is termed *drift-space bunching*, while that in the reflex klystron is known as *reflector bunching*. If both drift-space and reflector bunching are attempted in the same tube, the one counteracts the other and debunching results.

The Applegate diagram, illustrating the bunching action in the reflex klystron, is shown in Fig. 20. The space-time curve for each electron is a parabola. An analogy of the reflex-klystron bunching may be obtained by throwing a number of balls vertically into the air at equally spaced time intervals, each successive ball being thrown with a lower initial
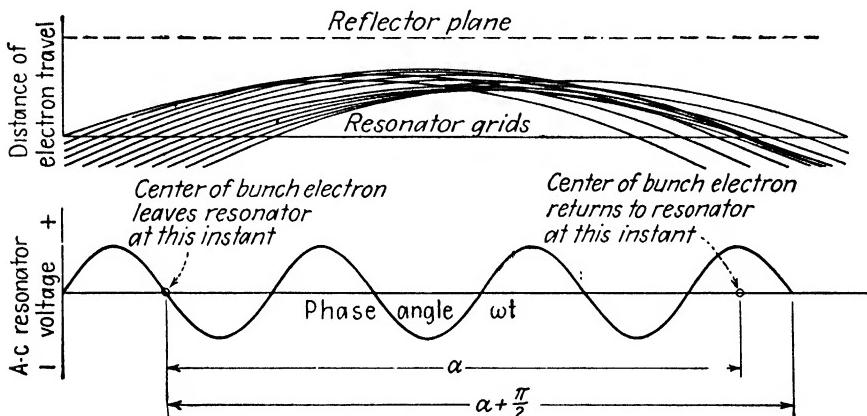
FIG. 20.—Electron bunching in the reflex klystron.

velocity. If the timing and initial velocities are properly chosen, all of the balls return to earth simultaneously. The space-time plot for the balls is then similar to that shown for the reflex klystron in Fig. 20.

The single-resonator feature of the reflex klystron greatly simplifies the tuning operation as compared with the double-resonator klystron. Oscillations may usually be obtained by merely adjusting the accelerating and reflector voltages. The frequency may be varied readily by changing the spacing between the grids of the resonator by means of a suitable tuning mechanism. A small variation in frequency may be obtained by varying either the accelerating voltage V_0 or the reflector voltage V_R. The frequency variation obtainable by varying the reflector voltage without retuning the resonator is of the order of 10 to 20 megacycles in a 3,000-megacycle oscillator.

The maximum theoretical efficiency of the reflex klystron is less than that of the double-resonator klystron. The reason for this is that in the reflex klystron a single resonator combines the functions of bunching the electrons and extracting energy from the returning bunches of electrons. The optimum value of the a-c resonator voltage, therefore, is a compromise

between the value required for optimum bunching and that required for maximum power transfer from the bunched electrons to the resonator field. Although the reflex klystron operates at a lower efficiency than the double-resonator klystron, it offers the attractive advantages of single-resonator tuning, simplicity of adjustment and operation, and compactness.

6.13. Analysis of the Reflex Klystron Oscillator.[1-3]—An expression relating the time of departure of electrons from the resonator grids and the return arrival time will be derived for the reflex klystron oscillator. As in the case of the double-resonator klystron, this will establish a relation-
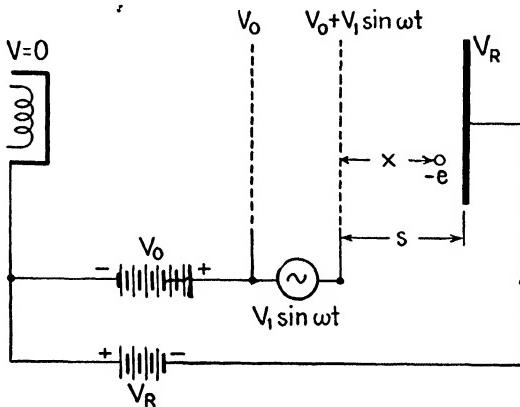
FIG. 21.—Diagram for the analysis of the reflex klystron.

ship between the various electrode potentials which must be satisfied if oscillating conditions are obtained. The derivations of the power output and efficiency equations are similar to those of the double-resonator klystron and are given as a problem.

Figure 21 shows the potentials of the various electrodes of the reflex klystron while oscillating. The cathode is taken as the zero-potential electrode and an a-c potential difference $V_1 \sin \omega t$ is assumed to exist between the grids. The potentials of grids 0 and 1 (with respect to the cathode) are assumed to be V_0 and $V_0 + V_1 \sin \omega t$, respectively, while that of the reflector electrode is V_R (a negative potential).

Referring to Fig. 21, we find that the electrons enter grid 0 with a velocity

$$v_0 = \sqrt{\frac{2V_0 e}{m}} \tag{1}$$

[1] PIERCE, J. R., Reflex Oscillators, *Proc. I.R.E.*, vol. 33, pp. 112–118; February, 1945.

[2] GINZTON, E. L., and A. E. HARRISON, Reflex Klystron Oscillators, *Proc. I.R.E.*, vol. 34, pp. 97–113P; March, 1946.

[3] HARRISON, A. E., Kinematics of Reflection Oscillators *J. Applied Phys.*, vol. 15, pp. 709–711; October, 1944.

As in Eq. (6.06-3) for the double-resonator klystron, the electrons leave grid 1 with a velocity

$$v_1 = v_0 \sqrt{1 + \frac{V_1}{V_0} \sin \omega t} \tag{2}$$

The potential difference between grid 1 and the reflector electrode is $V_R - (V_0 + V_1 \sin \omega t)$. We now assume that $V_1 \sin \omega t$ is small in comparison with the other terms and write the potential difference as $V_R - V_0$ (where V_R is usually a negative potential). If we assume a uniform field between grid 1 and the reflector electrode, the electric intensity is $E = -(V_R - V_0)/s$, where s is the distance between grid 1 and the reflector.

The force experienced by the electron in the reflecting region is $-eE = e[(V_R - V_0)/s]$. Equating force to mass times acceleration, we obtain the equation of electron acceleration,

$$\frac{d^2x}{dt^2} = \frac{e(V_R - V_0)}{ms} \tag{3}$$

where distance s is measured from grid 1. Two integrations successively give the velocity and displacement equations. The electron leaves grid 1 ($x = 0$) at time $t = t_1$ and with a velocity v_1. The electron displacement equation, with the constants evaluated for the given boundary conditions, then becomes

$$x = \frac{e(V_R - V_0)}{2ms} (t - t_1)^2 + v_1(t - t_1) \tag{4}$$

The electron is decelerated as it approaches the negative reflector electrode, reverses its direction of travel, and starts back toward the resonator grids. Letting t_2 represent the return arrival time of the electron, we may write Eq. (4) for the instant of arrival by substituting $t = t_2$ and $x = 0$. The total time of transit of the electron in the reflecting space $t_2 - t_1$ is therefore

$$t_2 - t_1 = -\frac{2msv_1}{e(V_R - V_0)} \tag{5}$$

The phase angle of departure of the electron, with respect to the buncher voltage, is ωt_1 and the phase angle of arrival is ωt_2. Multiplying Eq. (5) by ω and substituting the value of v_1 from Eq. (2), we obtain the arrival phase angle

$$\omega t_2 = \omega t_1 - \frac{2\omega msv_0}{e(V_R - V_0)} \left(1 + \frac{V_1}{V_0} \sin \omega t_1 \right)^{1/2} \tag{6}$$
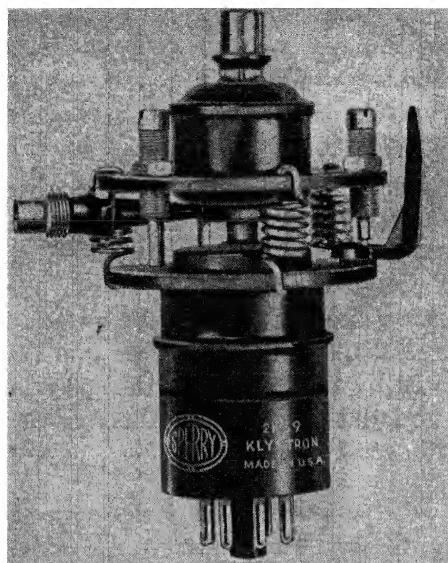
Fig. 22.—Sperry 2K39 reflex klystron. ` (*Courtesy of the Sperry Gyroscope Company.*)
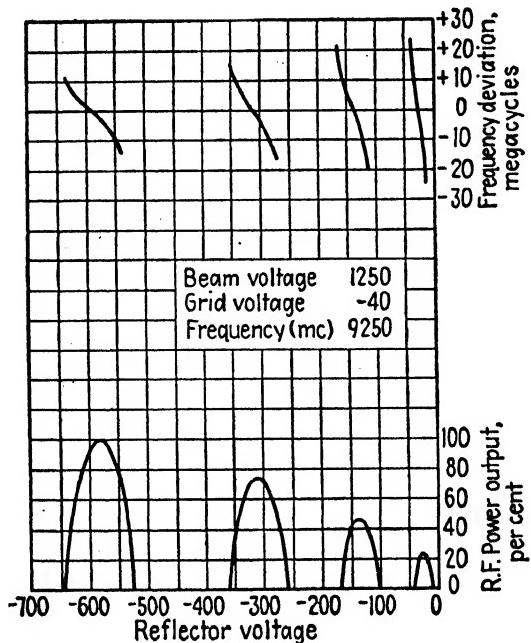
Fig. 23.—Power output and frequency deviation of a reflex klystron as a function of reflector voltage

Equation (6) is analogous to Eq. (6.06-5) for the double-resonator klystron. Following this example, we set

$$\alpha' = -\frac{2\omega m s v_0}{e(V_R - V_0)} \tag{7}$$

where α' is the round-trip transit angle of the center-of-the-bunch electron. Expanding the bracketed term in Eq. (6) using the binomial theorem and taking the first two terms of the series, we obtain

$$\omega t_2 = \omega t_1 + \alpha'\left(1 + \frac{V_1}{2V_0}\sin \omega t_1\right) \tag{8}$$

The bunching parameter is given by

$$x' = \frac{\alpha' V_1}{2V_0} \tag{9}$$

The values of α' and the bunching parameter $x' = \alpha' V_1/2V_0$ required for maximum power output and maximum efficiency may be found by methods similar to those used for the double-resonator klystron.

The value of α' required for sustained oscillation may be found by applying the criterion of oscillation described for the double-resonator klystron. In Fig. 20, the angle α' is found to be the phase angle between the zero of the resonator alternating potential (when the potential is changing from acceleration to deceleration) and the positive peak of the resonator voltage. The phase difference between two zeros of resonator voltage is $\alpha' + (\pi/2)$; therefore, we have $\alpha' + (\pi/2) = 2\pi n$. The relationship between accelerating voltage and reflector voltage required for oscillation is found by inserting $\alpha' = 2\pi n - (\pi/2)$ and v_0 from Eq. (1) into (7), yielding

$$\frac{V_0}{(V_R - V_0)^2} = \frac{[2\pi n - (\pi/2)]^2 e}{8\omega^2 s^2 m} \tag{10}$$

There are a number of discrete values of $V_0/(V_R - V_0)^2$ which produce oscillation. If V_R is the variable, the smaller values of n correspond to higher values of V_R.   The ratio V_1/V_0 is given by

$$\frac{V_1}{V_0} = \frac{2x'}{\alpha'} = \frac{2x'}{2\pi n - (\pi/2)} \tag{11}$$

The power output and efficiency for $\alpha = 2\pi n - (\pi/2)$ are

$$P_{ac} = I_0 V_1 J_1(x') = \frac{2V_0 I_0 x' J_1(x')}{2\pi n - \pi/2} \tag{12}$$

$$\eta = \frac{V_1}{V_0} J_1(x') = \frac{2x' J_1(x')}{2\pi n - \pi/2} \tag{13}$$

As n decreases, V_1/V_0 increases; hence the power output and efficiency both increase.

The efficiency is a maximum when the product $x'J_1(x')$ has its maximum value. The foregoing analysis was based upon certain assumptions which lead to a higher theoretical efficiency than is obtained by a more critical analysis. The maximum theoretical efficiencies of reflex klystrons range from 20 to 30 per cent.
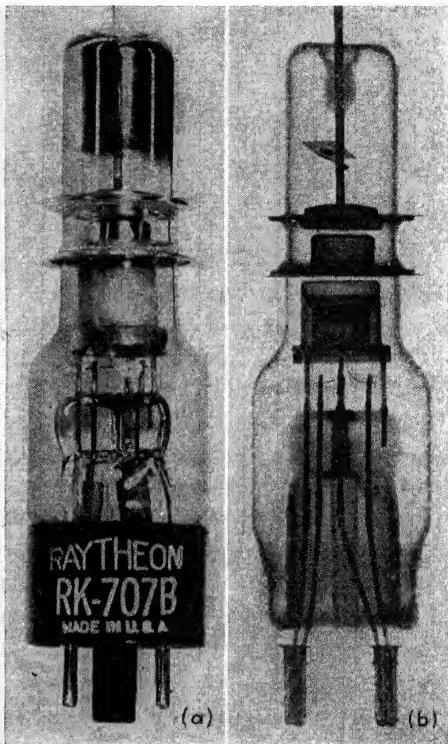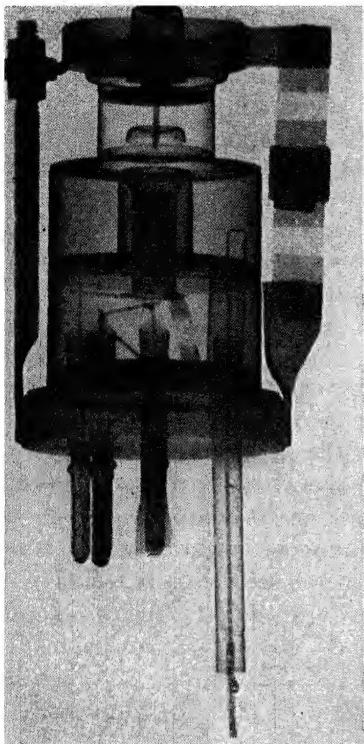
FIG. 24.—X-ray photograph of a Shepherd-Pierce reflex oscillator. (*Courtesy of the M.I.T. Radiation Laboratory.*)

FIG. 25.—McNally tube. (*Courtesy of the M.I.T. Radiation Laboratory.*)

6.14. Examples of Reflex Klystrons.—Two modifications of the reflex klystron are shown in Figs. 24 and 25. Figure 24 is an X-ray photograph of a Shepherd-Pierce tube which is used as an oscillator at wavelengths of 9 to 11 centimeters. This is an all-metal tube with a built-in resonator. At the top of the resonator is a thin corrugated diaphragm which permits changing the spacing between the grids, thereby tuning the oscillator. Tuning is accomplished by means of a bowed strut at the right-hand side of the tube which is clamped to the top of the tube. Adjustment of this

strut alters the spacing between grids. The coaxial output lead is coupled to the resonator by means of a small coupling loop which is clearly shown in the photograph. This tube has an output rating of 100 milliwatts and can be tuned over a range of 20 megacycles by varying the potential of the reflector electrode.

Figures 25a and 25b are photographs of a McNally tube. This tube is a reflex klystron which is used in conjunction with an external resonator (not shown in the illustration). The resonator is clamped to the grid disks and is usually tuned by screw plugs. This tube has a rating of 75 milliwatts at wavelengths of 8 to 12 centimeters.

In recent years considerable effort has been devoted to the development of reflex klystrons which can be tuned over a broad band of frequencies.[1-3] For practical reasons, it has been found desirable to build the tube and resonator as separate units. The McNally type of tube is used and the resonator usually consists of a coaxial line which contains an adjustable short-circuiting piston at one end for tuning. The line may operate either as a $\frac{1}{4}$ wavelength or a $\frac{3}{4}$ wavelength resonant line, with a voltage maximum appearing at the klystron grids. Power output is obtained by coupling either a probe or a loop to the resonant line. Tuning is accomplished by the simultaneous adjustment of the reflector voltage and the length of line.

In Eq. (6.13-10) it was shown that a number of different modes of oscillation can exist in a reflex klystron, the various modes corresponding to different values of the integer n. In general, the power output increases with decreasing values of n, but the stability of the oscillator under variable load conditions becomes poorer. If too large a value of n is used, the frequency of the oscillator is seriously affected by small changes in V_R. It has been found that values of n from 1 to 5 offer the most stable performance, with $n = 3$ being preferred. The mode of oscillation is determined, in part, by the reflector voltage. Consequently, in attempting to obtain continuous tuning by varying V_R and the length of line, it is found that mode jumping may occur, resulting in an abrupt change in frequency and power output. For example, if the resonant line was operating as a $\frac{1}{4}$ wavelength resonant line, the new frequency may be such that the line operates as a $\frac{3}{4}$ wavelength line. This tendency of mode jumping makes it difficult to obtain continuous tuning over a large range of frequencies. Despite this limitation, however, reflex klystrons have been constructed with tuning

[1] NELSON, R. B., Methods of Tuning Multiple-Cavity Magnetrons, *Phys. Rev.*, vol. 70, p. 118; July, 1946.

[2] KEARNEY, J. W., Design of Wide-Range Coaxial-Cavity Oscillators Using Reflex Klystron Tubes in the 1000 to 11,000 Megacycle Frequency Region, *Proc. National Electronics Conference*, vol. 2, pp. 624–636; October, 1946.

[3] CLARK, J. W., and A. L. SAMUEL, A Wide-Tuning-Range Microwave Oscillator Tube, *Proc. I.R.E.*, vol. 35, pp. 81–87; January, 1947.

ranges exceeding 2 to 1 in the range of frequencies from 1,000 to 11,000 megacycles.

Another recent development [1] consists of a velocity-modulated tube in which the electron bunches oscillate back and forth, passing through the resonator grids during successive cycles of oscillation in a manner similar to that of electron oscillation in the positive-grid oscillator. A mathematical analysis of this type of tube has shown that the electrons will be bunched at the center of the resonator grids during successive cycles of
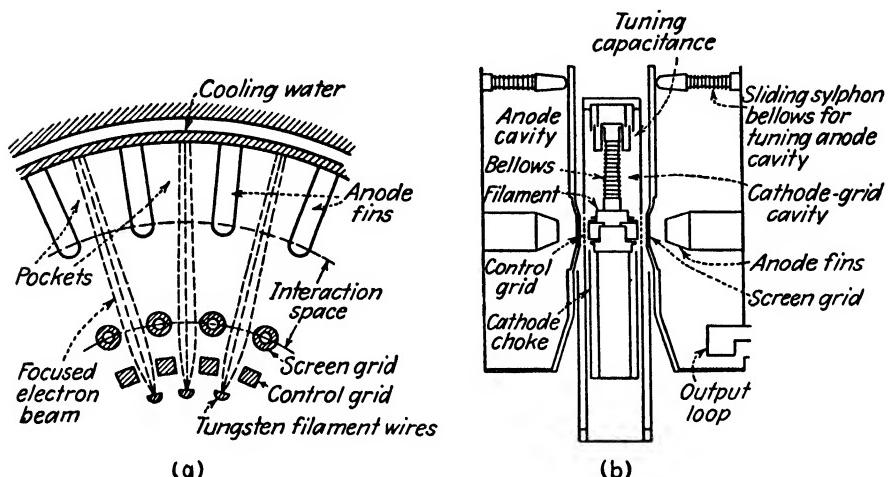
FIG. 26.—Sectional view of the resnatron.

oscillation if the d-c potential has a parabolic distribution on either side of the resonator grids. Efficiencies exceeding 40 per cent have been obtained with this type of tube.

6.15. The Resnatron. [2, 3]—The remarkable performance obtained with resnatron tubes indicates that this type of tube holds considerable promise for the future. At present, tubes have been built which deliver continuous power outputs of 85 kilowatts with anode efficiencies of 60 to 70 per cent, operating in the frequency range from 350 to 650 megacycles. It appears quite likely that the frequency range of this type of tube can be extended farther into the microwave-frequency spectrum.

The resnatron, shown in Fig. 26, is a cylindrical tetrode with an input resonator between the cathode and control grid and an output resonator

[1] COETERIER, F., The Multireflection Tube—A New Oscillator for Very Short Waves, *Philips Tech. Rev.*, vol. 8, pp. 257–266; September, 1946.

[2] SALISBURY, W. W., The Resnatron, *Electronics*, vol. 19, pp. 92–97; February, 1946.

[3] DOW, W. G., and H. W. WELCH, The Generation of Ultra-High Frequency Power at the Fifty Kilowatt Level, *Proc. National Electronics Conference*, vol. 2, pp. 603–614; October 1946.

between the screen grid and anode. The input and output resonators are coupled together by an external coaxial line when the tube is used as an oscillator. The control grid is biased beyond cutoff so that the operation of the tube is similar to that of a class C triode oscillator. However, the resnatron differs from the triode oscillator in two important respects. In the triode tube, the electrons are simultaneously accelerated by the d-c field and retarded by the a-c field. This results in relatively low electron velocities and the electron transit times may therefore be excessively large at microwave frequencies. In the resnatron, the electrons are accelerated by the d-c field between the control grid and screen grid and consequently attain a high velocity before entering the a-c field between the screen grid and plate. This results in an appreciable reduction in electron transit time through the a-c field, making it possible to obtain greater power output and higher efficiency. Also, by coupling the input and output resonators together by means of an external coaxial line, it is possible to vary the length of this line and thereby introduce any desired amount of phase shift between the two resonators to compensate for electron transit-time delay. In this way the electron bunches are made to traverse the a-c field between screen grid and plate at the optimum phase angle for maximum power output.

The cathode of the resnatron of Fig. 26 is the dark, shaded portion. This consists of 24 pure tungsten filament wires, each about 1 inch long, which are hard soldered to two copper rings. Quarter-wavelength chokes are placed on the cathode rod below the cathode to prevent radio-frequency loss. The control grid is a piece of copper tubing containing a number of longitudinal slots, one for each filament wire. The screen grid consists of a squirrel-cage grid of copper tubing. This is connected to the anode structure; hence the screen grid and anode are at the same d-c potential. The anode is a copper annular ring containing fins which project radially inward. By properly shaping the cathode, control grid, and screen grid electrodes, the electron beam may be focused in such a way as to minimize screen-grid current, thereby preventing high screen-grid dissipation.

Both the input and output resonators consist of coaxial lines which are effectively ¾ of a wavelength long, with the maximum a-c voltage points located at the tube electrodes. The input resonator is tuned by means of a lumped capacitance, shown at the top of the resonator in Fig. 26. The output resonator is tuned by means of a plunger and an expanding bellows, which is shown at the top of the output resonator. The output loop is coupled to a wave guide which has cross-sectional dimensions 6 by 15 inches (for 650 megacycles). Vacuum is maintained by an exhaust pump.

For typical operating conditions, the anode potential is 17½ kilovolts, control-grid potential −2,500 volts, filament current 1,800 amperes, and grid current 1½ amperes.

6.16. The Traveling-wave Tube.[1-3]—The traveling-wave tube is a new type of tube which has shown considerable promise as a broad-band amplifier. As shown in Fig. 27, the tube contains a closely wound wire helix through which an electron beam travels. The signal to be amplified is impressed upon the helix at the input end of the tube and takes the form of a wave traveling along the helix. The interaction of the electron stream and the traveling wave results in an amplification of the signal as it progresses along the helix. The amplified signal is then removed at the output end of the helix.

If the length of wire in the helix is n times as long as the helix, the velocity of the traveling wave is approximately $v = v_c/n$, where v_c is the velocity of light. The electrons enter the helix with a velocity slightly greater than
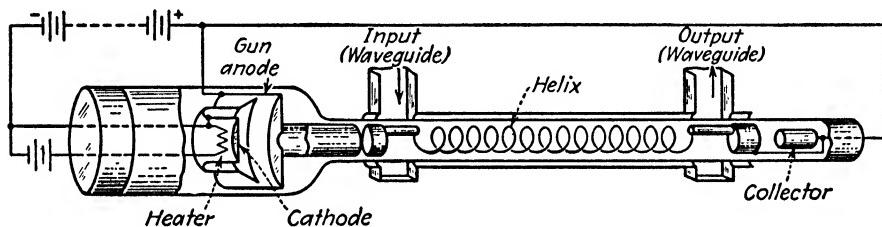
FIG. 27.—Traveling-wave tube.

that of the traveling wave and tend to become bunched as they travel through the helix. Those electrons which are retarded by the a-c field of the traveling wave slow down, while those electrons which are accelerated by the a-c field speed up and overtake the preceding slower electrons. In this way the electrons tend to bunch in the regions of the retarding a-c field. This results in energy transfer from the electrons to the traveling wave, thereby building up the amplitude of the wave. The increase in amplitude of the traveling wave is somewhat analogous to the building up of water waves by a wind blowing past them.

A longitudinal magnetic field (not shown in Fig. 27) is used to focus the electrons, thereby compelling them to travel in a direction parallel to the axis of the helix. Electrons that move parallel to the magnetic field are not affected by the magnetic field, but electrons which have a component of motion across the field experience a torque which tends to force them to move in the direction of the magnetic field.

[1] PIERCE, J. R., Traveling-Wave Tubes, *Proc. I.R.E.*, vol. 35, pp. 108–111; February, 1947.

[2] PIERCE, J. R., Theory of the Beam-Type Traveling-Wave Tube, *Proc. I.R.E.*, vol. 35, pp. 111–123; February, 1947.

[3] KOMPFNER, R., The Traveling-Wave Tube as an Amplifier of Microwaves, *Proc. I.R.E.*, vol. 35, pp. 124–127; February, 1947.

The traveling-wave tube shown in Fig. 27 is untuned and therefore operates as a broad-band amplifier. Bandwidths as high as 800 megacycles have been realized with this type of tube. This is of the order of 80 times the bandwidth possible in a single stage video amplifier using a conventional pentode tube.

PROBLEMS

1. A positive-grid oscillator, connected as shown in Fig. 1, has a d-c potential difference of 100 volts between grid and plate (or cathode) and a peak a-c potential of 75 volts. The distance between the cathode and plate is 1 cm and the grid is midway between the cathode and plate. Compute the frequencies of oscillation of electrons which leave the grid plane at values of $\phi = 0, \pi/2, \pi,$ and $3\pi/2$. What would be the approximate frequency of electrical oscillation as a Gill-Morrell oscillator? What would be the approximate frequency of electrical oscillation as a Barkhausen-Kurz oscillator?

2. Derive expressions for the power input, power output, and conversion efficiency of an electron moving in the grid-plate region of the positive-grid oscillator shown in Fig. 1. The grid to plate (or cathode) voltage is $V = -V_0 - V_1 \sin \omega t$ and the electron leaves the grid plane at the phase angle ϕ. Assume that the alternating potential is small and therefore that the electron velocity is determined by the d-c potential only.

3. A double-resonator klystron is tuned to a frequency of 3,000 megacycles. The dimensions of the klystron, as shown in Fig. 10, are $d_1 = d_2 = 0.1$ cm and $s = 2$ cm (see Fig. 10). The beam current is $I_0 = 25$ ma and the ratio of V_2/V_0 is 0.3. The two resonators are assumed to oscillate in time phase.

   (a) Compute the d-c accelerating voltage, buncher voltage, and catcher voltage required for maximum power output for integer values of n from 1 to 5.

4. Compute the power output, power input, and maximum theoretical efficiency of the klystron given in Prob. 3, using values of $n = 2$ and 3. In computing the efficiency, take into consideration the reduction in efficiency caused by the electron transit time delay in passing through the catcher resonator.

5. Obtain a mathematical expression for the curves of Fig. 17 by differentiating Eq. (6.10-1) and substituting this into Eq. (6.10-4) to eliminate ωt_2. What values of ωt_1 and ωt_2 correspond to infinite values of q_{r2}/q_{r0} for $x = 1.0$ and $x = 1.84$?

6. A reflex klystron is tuned to a frequency of 5,000 megacycles. The distance s is 0.5 cm (see Fig. 19). What values of $V_0/(V_R - V_0)^2$ are required for oscillation corresponding to $n = 1$ to 5? Specify values of V_0 and V_R which will produce oscillation.

7. Derive the equations for the power output, power input, and efficiency of the reflex klystron.

8. Using Eq. (6.13-10) derive an expression for df/dV_R for the reflex klystron. Assume that $V_0 = 350$ volts, $s = 0.5$ cm, $f = 5,000$ megacycles, and $n = 4$. Compute the frequency change for a 1-volt change in reflector potential.

CHAPTER 7

MAGNETRON OSCILLATORS

During the early stages of the war, the urgent quest for a suitable microwave generator, capable of delivering large power output at high efficiencies under pulsed conditions, for use in high-definition radar, led to the development of the multicavity magnetron. The British brought to the United States an early model of the multicavity magnetron in the fall of 1940 and demonstrated its potentialities as a source of microwave power. Thereafter its development proceeded rapidly under the stimulus of an intensive wartime research program.

7.01. Description of Multicavity Magnetrons.—A cutaway view of a 10-centimeter multicavity magnetron is shown in Fig. 1. This contains an indirectly heated oxide-coated cathode and a laminated copper anode. The anode block is bored or broached to provide for eight identical cylindrical resonators. A pickup loop in one of the resonators is connected to the external circuit by means of a vacuum-sealed coaxial line to provide the power-output connection. The magnetron of Fig. 1 contains disk-shaped fins for forced-air cooling.

Figure 2 shows a 10-centimeter magnetron and a 3.2-centimeter magnetron, complete with alnico permanent magnets. The output of the small magnetron is coupled to a wave guide through a vacuum-sealed dielectric window.

In the operation of the magnetron, a high d-c potential is applied between cathode and anode, setting up a radial electric field. An axial magnetic field is provided by either a permanent magnet or an electromagnet. Electrons emitted from the cathode experience a force directed radially outward due to the d-c electric field and a force perpendicular to their instantaneous direction of motion due to the magnetic field. The combined forces cause the electrons to take a spiral path, the radius of curvature of which decreases with increasing magnetic field strength. For a given value of d-c anode potential, there is a critical value of magnetic field strength which causes the electrons to just graze the anode. This value of magnetic field strength is known as the cutoff value. If the magnetic field strength exceeds the cutoff value, the curvature of the electron path is such that the electrons miss the anode and spiral back to the cathode. Magnetic field strengths of the order of one to two times cutoff are normally required for oscillation.

The various types of oscillation of a magnetron will be considered in detail later in the chapter.   However, the following brief description is intended to give an over-all picture of the mechanism of the traveling-wave type of oscillation.

When the magnetron is functioning as an oscillator, the electrical oscillation of the resonators sets up an a-c electric field across the resonator
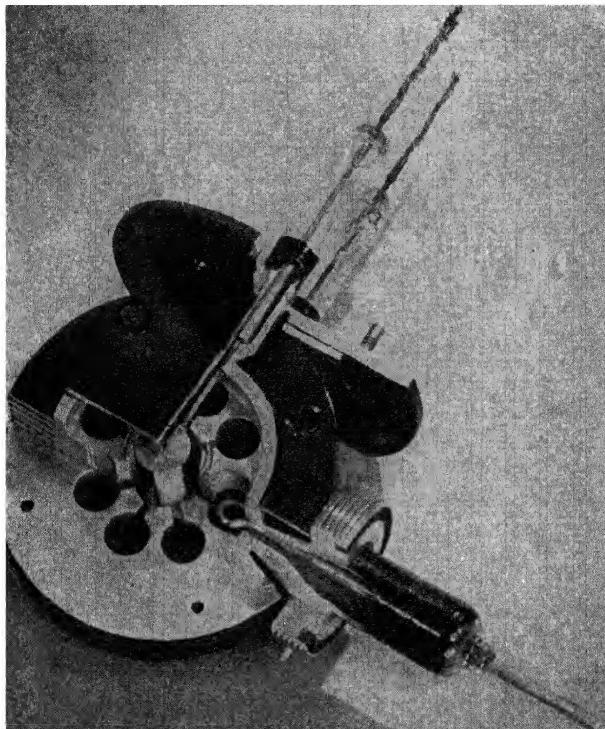
FIG. 1.—Cutaway view of a 10-centimeter cavity magnetron.   (*Courtesy of the M.I.* *Radiation Laboratory.*)

gaps.   In the multianode magnetron, such as that shown in Fig. 1, the a-c field in the interaction space (the space between the cathode and anode) is largely tangential.   In general, there is a phase difference between the oscillations of successive resonators which produces a rotating a-c field. The d-c anode potential and the magnetic field strength are adjusted so that the whirling cloud of electrons rotates in synchronism with either the fundamental or a submultiple component of the rotating a-c field.

Part of the electrons rotate in such a phase that they are accelerated by the a-c field, while other electrons rotate in such a phase as to be retarded

by the a-c field. Those electrons which are accelerated by the a-c field experience an increase in velocity and a consequent increase in torque resulting from their motion in the magnetic field. This causes these electrons to have their paths so altered that they spiral back to the cathode and they are thereby withdrawn from the interaction space. The electrons remaining in the interaction space are those which, on the average, are retarded by the a-c field. These electrons give energy to the a-c field thereby contributing to the power output of the magnetron. As a result of this
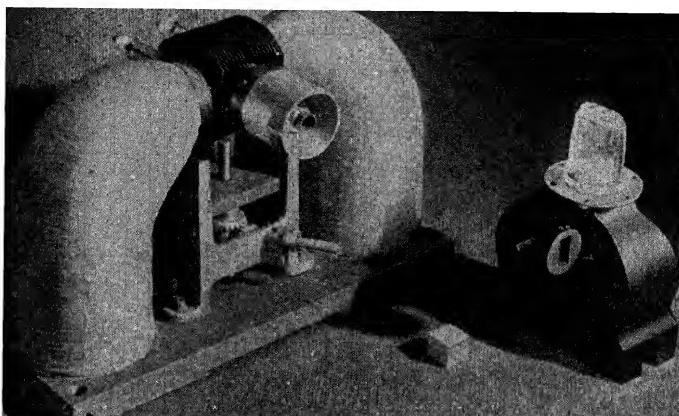
FIG. 2.—Ten- and three-centimeter magnetrons with magnets.   (*Courtesy of the Western Electric Company.*)

selection process, the electrons form into groups resembling the spokes of a wheel, with the entire spokelike formation rotating at synchronous speed with respect to a component of the a-c field.

The foregoing description of the operation of the magnetron applies to the traveling-wave type of oscillation. The magnetron may also operate as a negative-resistance oscillator or as a cyclotron-frequency oscillator. These types of oscillation will be described in detail later.

7.02. Magnetrons as Pulsed Oscillators.—The magnetron is particularly well suited to pulsed operation, such as is required in radar systems and pulse-time modulation. In a pulsed system, a high d-c anode potential is applied to the magnetron for an extremely short interval of time, with a relatively long time interval between pulses during which the tube is inoperative. This makes it possible to obtain very high values of peak power output and still remain within safe limits of average power output and plate dissipation for the tube.

A water-cooled magnetron, capable of delivering 2.5 megawatts peak power output at a wavelength of 10 centimeters under pulsed conditions,

is shown in Fig. 3. A typical pulse cycle for this tube consists of a pulse duration of 1 microsecond, with 1,000 pulses per second. If the peak power output is 2.5 megawatts, the average power would be 2.5 kilowatts. The cathode current during the pulse interval is 140 amperes and the peak anode voltage is 50 kilovolts.

Pulsed operation of a magnetron imposes extremely severe requirements upon the cathode. Current densities of the order of 50 amperes per square centimeter are sometimes required. Experiments have shown that the
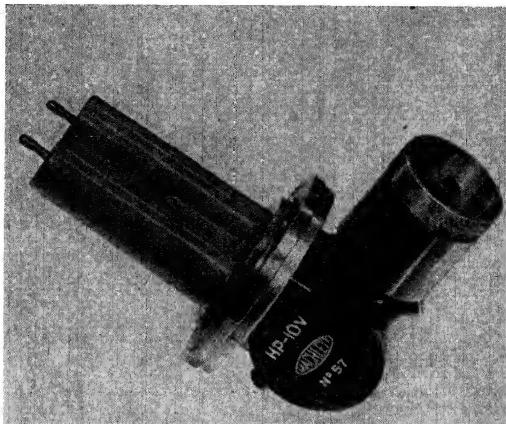
FIG. 3.—High-power cavity magnetron. (*Courtesy of the M.I.T. Radiation Laboratory.*)

cathode current under pulsed conditions may be many times the current obtainable when the tube is not oscillating. The reason for this very large cathode current is not entirely clear. One possible explanation is that part of the electrons (those rotating in an unfavorable phase with respect to the a-c field) return to the cathode with relatively high kinetic energy and consequently splash out secondary electrons. Since a portion of the secondary electrons emitted from the cathode have an unfavorable phase with respect to the a-c field, these electrons return to the cathode and splash out other secondary electrons, resulting in a cumulative bombardment of the cathode which builds up the emission to very large values. Magnetrons designed for short-wavelength operation sometimes depend entirely upon this bombardment of the cathode by returning electrons to heat the cathode, there being no additional heater power supplied. Occasionally the back bombardment becomes excessive, resulting in overheating and damage to the cathode.

Another possible explanation for the high cathode emission is that the work function of the cathode may be lowered due to relatively large fields at the cathode surface, ionic conduction, or electrolytic conduction. A

certain amount of the increased current can be attributed to the space-charge cloud which has accumulated around the cathode during the quiescent period.

Cathodes are usually constructed of a nickel cylinder or a nickel wire mesh containing an oxide coating. Special oxide coatings have been developed for the short-wavelength magnetrons to overcome the limitation of the power output resulting from the small size of the cathode.
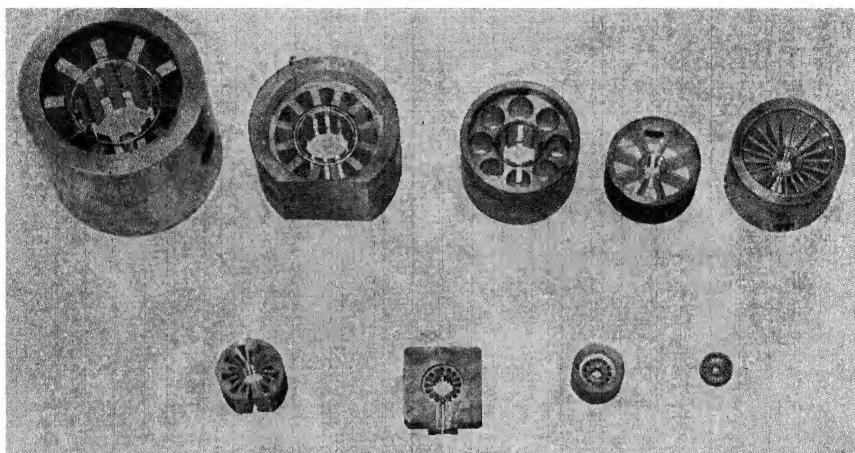
Fig. 4.—Anodes of 10-centimeter and 3-centimeter magnetrons. *(Courtesy of the M.I.T. Radiation Laboratory.)*

7.03. Electron Motion in Uniform Magnetic Fields.—If a charge q is projected into a uniform magnetic field with an entering velocity v, it experiences a force in a direction mutually perpendicular to its instantaneous direction of motion and to the direction of the magnetic field. This is known as the *Lorentz force* and is given by

$$f = qvB \sin \theta \tag{1}$$

where B is the magnetic flux density and θ is the angle between B and v. In mks units, B is in webers per square meter.

The Lorentz force may also be expressed in vector notation as a cross product, thus

$$\bar{f} = q\bar{v} \times \bar{B} \tag{2}$$

The direction of the vector force $\bar{f}$ is found by applying the right-hand rule described in Sec. 2.01.

Figure 5 shows the motion of an electron in a uniform magnetic field when the magnetic field is directed into the page. The force on an elec-

tron is directed opposite to that of a positive charge, as is evident if we substitute $q = -e$ in Eq. (2).

If the electron moves in a plane perpendicular to the magnetic field, the path is a circle. The Lorentz force, directed radially inward, is equal and
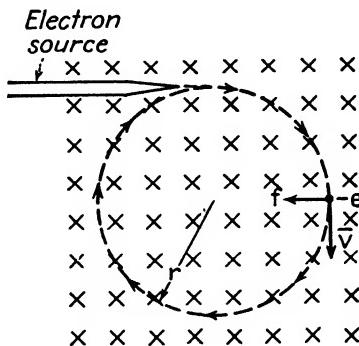
FIG. 5.—Electron motion in a uniform magnetic field.

opposite to the centrifugal force mv^2/r, where r is the radius of the path. Equating the magnitudes of the Lorentz and centrifugal forces and setting $\sin \theta = 1$ in Eq. (1), we obtain

$$evB = \frac{mv^2}{r} \qquad (3)$$

The equations for the radius of the path, angular velocity, and period of rotation, as obtained from Eq. (3), are as follows:

$$r = \frac{mv}{eB} \qquad (4)$$

$$\omega = \frac{v}{r} = \frac{eB}{m} \qquad (5)$$

$$T = \frac{2\pi}{\omega} = \frac{2\pi m}{eB} \qquad (6)$$

The radius of the path varies directly with electron velocity and inversely as the magnetic field strength. The kinetic energy and the magnitude of the velocity of a charge moving in a stationary magnetic field are constant. In general, a stationary magnetic field can neither give energy to nor take energy from a charge moving in the field.

7.04. Electron Motion in the Parallel-plane Magnetron.—Let us now consider the electron trajectories and cutoff criterion for the idealized parallel-plane magnetron shown in Fig. 6. The electric intensity is directed in the negative z direction, and the magnetic intensity is assumed to be y directed.

The vector force on a charge in combined electric and magnetic fields is obtained by adding Eqs. (2.06-1) and (7.03-2), yielding

$$\bar{f} = -e(\bar{E} + \bar{v} \times \bar{B}) \tag{1}$$

This force may be expressed in terms of its scalar components. The z-directed force includes a component due to the electric field and a Lorentz force due to the velocity component v_x; thus $f_z = -e(E + v_x B)$. The x-
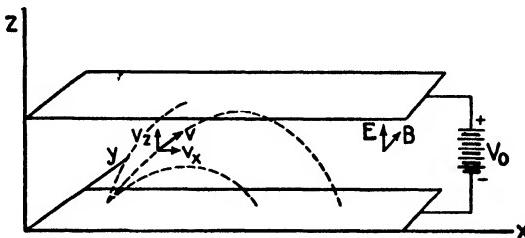
FIG. 6.—Parallel-plane magnetron.

directed force consists only of a Lorentz force due to velocity v_z, or $f_x = ev_z B$. Equating the force components to mass times acceleration, we obtain

$$ev_z B = ma_x \tag{2}$$

$$-e(E + v_x B) = ma_z \tag{3}$$

These equations may be expressed in terms of velocity, by substituting the relationships $a_x = dv_x/dt$ and $a_z = dv_z/dt$ into Eqs. (2) and (3), and, with the additional substitution of $E = -V_0/d$, we obtain

$$\frac{dv_x}{dt} = \omega_e v_z \tag{4}$$

$$\frac{dv_z}{dt} = \frac{eV_0}{md} - \omega_e v_x \tag{5}$$

where

$$\omega_e = \frac{eB}{m} \tag{6}$$

Comparing Eq. (6) with (7.03-5), we find that the expression for ω_e is the same as that for the angular velocity in a pure magnetic field.

To obtain an explicit equation in one variable, differentiate Eq. (5) with respect to time and substitute Eq. (4) for dv_x/dt. Since the potential V_0 is a constant, we have

$$\frac{d^2 v_z}{dt^2} = -\omega_e^2 v_z \tag{7}$$

A solution of this equation is $v_z = C_1 \sin \omega_e t + C_2 \cos \omega_e t$. To evaluate the constants, assume that the electron leaves the cathode at zero time and with zero initial velocity. Substituting these boundary conditions into the equation for v_z we find that $C_2 = 0$. Thus, there remains

$$v_z = C_1 \sin \omega_e t \tag{8}$$

To evaluate the constant C_1, differentiate Eq. (8) with respect to time, yielding $dv_z/dt = C_1 \omega_e \cos \omega_e t$. Now equate this to Eq. (5) and write the resulting equation for zero time, remembering that when $t = 0$, we also have $v_x = 0$. The constant C_1 then becomes $C_1 = eV_0/\omega_e md$. The z component of velocity is obtained by substituting the constant C_1 into Eq. (8). The x component of velocity is found by substituting v_z into Eq. (5). The velocity components then become

$$v_z = \frac{eV_0}{\omega_e md} \sin \omega_e t \tag{9}$$

$$v_x = \frac{eV_0}{\omega_e md} (1 - \cos \omega_e t) \tag{10}$$

Inserting $v_z = dz/dt$ and $v_x = dx/dt$ into the above expressions and integrating, we obtain the electron-displacement equations. The integration constants are evaluated by assuming that the electron leaves the origin at zero time; hence when $t = 0$ we have $x = z = 0$, and

$$z = \frac{eV_0}{\omega_e^2 md} (1 - \cos \omega_e t) \tag{11}$$

$$x = \frac{eV_0}{\omega_e^2 md} (\omega_e t - \sin \omega_e t) \tag{12}$$

The path taken by the electrons, as determined by Eqs. (11) and (12), is a cycloid. A cycloid is the locus of a point on the circumference of a circle which is rolling along a straight line. From Eq. (11), we find that the maximum displacement of the electron in the z direction occurs when $\omega_e t = \pi$. The maximum displacement is therefore

$$z_{\max} = \frac{2eV_0}{\omega_e^2 md} = \frac{2mV_0}{eB^2 d} \tag{13}$$

Cutoff occurs when the electron just grazes the anode, or when $z_{\max} = d$. Making this substitution in Eq. (13) and solving for the cutoff value of magnetic flux density, we obtain

$$B_c = \frac{1}{d} \sqrt{\frac{2mV_0}{e}} \tag{14}$$

Equation (11) shows that the electron leaves the cathode when $\omega_e t = 0$ and returns when $\omega_e t = 2\pi$. The transit time of the electron for one complete cycle is, therefore,

$$T = \frac{2\pi}{\omega_e} = \frac{2\pi m}{eB} \qquad (15)$$

The foregoing equations were derived for electrons moving in a plane perpendicular to the magnetic field. The relationships are more complicated if the electron has a component of velocity in the direction of the magnetic field. A parallel-plane magnetron, as such, would have little practical value, since successive oscillations of an electron would soon carry it out of bounds in the x direction.

7.05. Analysis of the Cylindrical-anode Magnetron.—We now consider the behavior of electrons in the cylindrical-anode magnetron under d-c operating conditions. In Fig. 7, the positive direction of the electric intensity is, by definition, radially outward (the actual electric intensity is radially inward and therefore is negative). The magnetic flux-density vec-
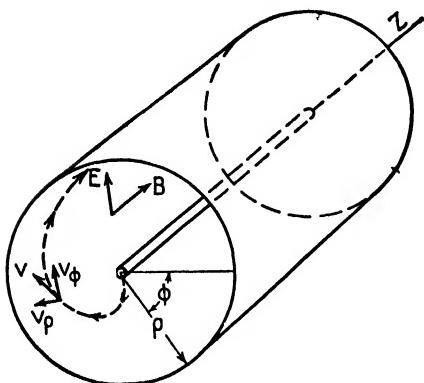
Fig. 7.—Diagram for the analysis of the cylindrical-anode magnetron.

tor is assumed to be in the z direction. Cylindrical coordinates ρ, ϕ, and z are used.

In formulating equations for rotational motion, it is convenient to use torque instead of force. A fundamental law of mechanics states that torque is equal to the time rate of change of angular momentum. The Lorentz force on an electron, due to its motion in a magnetic field, is $\bar{f} = -e(\bar{v} \times \bar{B})$, its direction being mutually perpendicular to $\bar{v}$ and $\bar{B}$. The ϕ component of this force is $f_\phi = ev_\rho B$, and the corresponding torque is $\rho f_\phi = \rho e v_\rho B$. To derive the angular momentum, we start with the moment of inertia of the electron $m\rho^2$ and multiply this by the angular velocity $d\phi/dt$ to obtain $m\rho^2(d\phi/dt)$. Equating torque to time rate of change of angular momentum, we get

$$eB\rho \frac{d\rho}{dt} = \frac{d}{dt}\left(m\rho^2 \frac{d\phi}{dt} \right) \qquad (1)$$

Multiplying both sides of Eq. (1) by dt and integrating, we have

$$\frac{eB\rho^2}{2} = m\rho^2 \frac{d\phi}{dt} + C_1 \qquad (2)$$

To evaluate the integration constant C_1 we recall that at the cathode ($\rho = a$) the angular velocity of the electron $d\phi/dt$ is zero. Equation (2) then yields the value of the constant $C_1 = eBa^2/2$. Equation (2) therefore gives the angular velocity as

$$\frac{d\phi}{dt} = \frac{eB}{2m}\left(1 - \frac{a^2}{\rho^2}\right)$$

$$= \omega_e\left(1 - \frac{a^2}{\rho^2}\right) \tag{3}$$

where, in the cylindrical magnetron, we let

$$\omega_e = \frac{eB}{2m} \tag{4}$$

According to Eq. (3), the angular velocity at the cathode ($\rho = a$) is zero. The angular velocity increases with radius, approaching an asymptotic value of $d\phi/dt = \omega_e$ at points such that $\rho \gg a$. The value of ω_e for the cylindrical magnetron is one half of the value for the parallel-plane magnetron.

We need an additional fundamental relationship relating the motion of the electron to the potential. An electron in motion in an electric field has a kinetic energy of $\frac{1}{2}mv^2$ and potential energy of $-eV$. The kinetic and potential energies are not altered by the presence of a magnetic field. If the fields are stationary, the sum of kinetic plus potential energy remains constant as the electron moves through the field. Assuming that the cathode is at zero potential and that the electron is emitted with zero velocity, the sum of kinetic plus potential energy at the cathode is zero. Consequently, at any other point in space, we have $\frac{1}{2}mv^2 - eV = 0$. The velocity has two components: a radial velocity v_ρ and a ϕ component of velocity v_ϕ. The resultant velocity is $v = \sqrt{v_\rho^2 + v_\phi^2}$. Substitution of v in the energy equation yields

$$eV = \frac{1}{2}m(v_\rho^2 + v_\phi^2)$$

$$= \frac{1}{2}m\left[\left(\frac{d\rho}{dt}\right)^2 + \rho^2\left(\frac{d\phi}{dt}\right)^2\right] \tag{5}$$

In order to solve for the cutoff conditions, substitute Eq. (3) into (5), yielding

$$eV = \frac{1}{2}m\left[\left(\frac{d\rho}{dt}\right)^2 + \omega_e^2\rho^2\left(1 - \frac{a^2}{\rho^2}\right)^2\right] \tag{6}$$

At the anode $\rho = b$ the potential is $V = V_0$ and Eq. (6) becomes

$$eV_0 = \frac{1}{2}m\left[\left(\frac{d\rho}{dt}\right)^2 + \omega_e^2 b^2\left(1 - \frac{a^2}{b^2}\right)^2\right] \tag{7}$$

Cutoff occurs when the radial velocity $d\rho/dt$ is zero at the anode. Thus, the condition for cutoff is obtained by substituting $B = B_c$, $d\rho/dt = 0$, and $\omega_e \doteq eB_c/2m$ in Eq. (7), yielding

$$V_0 = \frac{m\omega_e^2 b^2}{2e}\left(1 - \frac{a^2}{b^2}\right)^2 = \frac{eB_c^2 b^2}{8m}\left(1 - \frac{a^2}{b^2}\right)^2 \qquad (8)$$

$$B_c = \frac{1}{b[1 - (a^2/b^2)]}\sqrt{\frac{8mV_0}{e}} \qquad (9)$$

where B_c is the cutoff value of magnetic flux density. In most magnetrons, we have $b \gg a$ and Eq. (9) reduces approximately to

$$B_c = \frac{1}{b}\sqrt{\frac{8mV_0}{e}} = \frac{6.75 \times 10^{-6}}{b}\sqrt{V_0} \quad \text{(mks units)} \quad (10)$$

The expression for B_c given by Eq. (10) is similar to that of Eq. (7.04-14) for the parallel-plane magnetron. The foregoing derivations for the cylindrical magnetron are valid for any degree of space-charge density. An

B<Bc   B=Bc   B>Bc
(a)   (b)   (c)
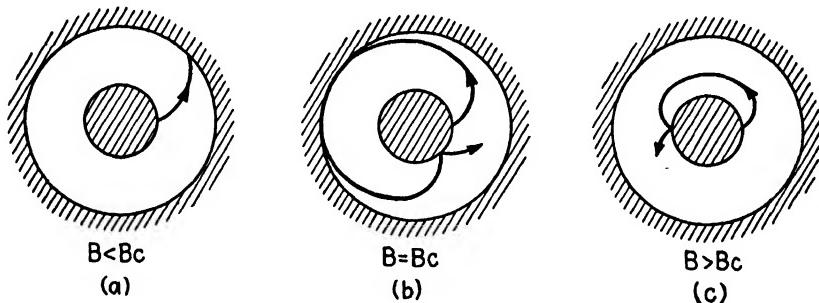
FIG. 8.—Electron paths in a cylindrical magnetron for various values of magnetic flux density.

electron, moving in a cylindrical magnetron under d-c operating conditions, describes approximately an epicycloidal path. An epicycloid is the locus of a point on the circumference of a circle which rolls along the circumference of another circle (the cathode). The electron paths for several different values of magnetic field strength are shown in Fig. 8. Cutoff condition is illustrated by Fig. 8b, and Fig. 8c corresponds to magnetic field strengths greatly exceeding cutoff.

Equation (9) expresses a critical value of magnetic flux density above which we would expect the current to drop abruptly to zero. However, measurements of plate current as a function of magnetic flux density show a more gradual decrease in plate current, as shown in Fig. 9. The gradual decrease in plate current in the vicinity of cutoff may be attributed to the

following causes: (1) electrons are emitted from the cathode with random emission velocities, and (2) electron collisions and space-charge field effects alter the individual electron velocities.
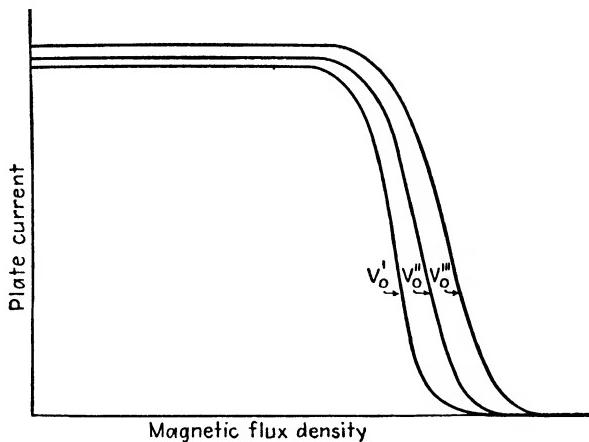
FIG. 9.—Plate current of a magnetron as a function of magnetic flux density and anode voltage

7.06. Negative-resistance Oscillation.—The negative-resistance magnetron utilizes a split-anode construction, usually having two anode segments. Since this type of magnetron is seldom used at frequencies exceed-
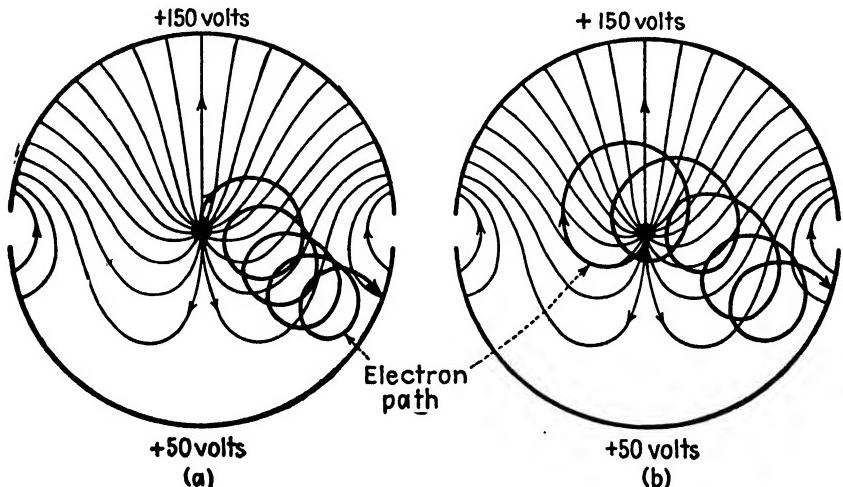
FIG. 10.—Electron paths in a split-anode magnetron having different values of d-c potential applied to the two anode segments.

ing 800 megacycles, the resonant circuit is usually mounted external to the tube. A common type of resonant circuit consists of a transmission line

having one end connected to the two anode segments and with an adjustable short-circuiting bar at the other end.

Oscillations occur in a negative-resistance oscillator by virtue of a static negative-resistance characteristic between anode voltage and anode current. Kilgore [1] showed that if the two anode segments have different values of d-c potential with respect to the cathode, within a certain range of potentials, the electron orbits are such that a majority of electrons travel to the least positive anode. This negative-resistance characteristic makes it possible to utilize the magnetron as a negative-resistance oscillator. Power outputs of several hundred watts at efficiencies as high as 50 to 60 per cent have been obtained by this means. In this type of oscillator there is no relationship between electron transit time and the period of electrical oscillation other than the requirement that transit time be small in comparison with the period of electrical oscillation.

In order to account for the negative-resistance characteristic, Kilgore took photographs of the electron paths in a magnetron. A small amount of gas was admitted to the tube so that ionization of the gas by the electron stream provided a luminous trace of the electron path. A shield, placed over the cathode, contained a small aperture in order to confine the emission to a single spot on the cathode surface. The electron paths shown in Fig. 10 were obtained in this manner. These show a tendency of the electrons to spiral out of the region of high potential and into the region of low potential, eventually terminating at the low-potential anode segment. Figure 11 shows the negative-resistance characteristic of the magnetron obtained in this manner.

7.07. Cyclotron-frequency Oscillation.—One type of oscillation of a magnetron has been found to have a period of electrical oscillation which is approximately equal to the electron transit time from cathode to the

Fig. 11.—Negative-resistance characteristic of the split-anode magnetron.

[1] Kilgore, G. R., Magnetron Oscillators for the Generation of Frequencies Between 300 and 600 Megacycles, *Proc. I.R.E.*, vol. 24, pp. 1140–1158; August, 1936.

vicinity of the anode and back to the cathode. This transit time is known as the *cyclotron period* of the electron; hence the associated oscillation is referred to as the *cyclotron-frequency* type. For this type, the product λH is a constant, specifically

$$\lambda H = 12,000 \qquad \text{(mks units)}$$

$$= 15,000 \qquad \text{(emu units)} \qquad (1)$$

Cyclotron-frequency oscillations may exist in either a single-anode or a split-anode magnetron. If a single-anode magnetron is used, the resonant circuit is connected between cathode and anode. The selection process,
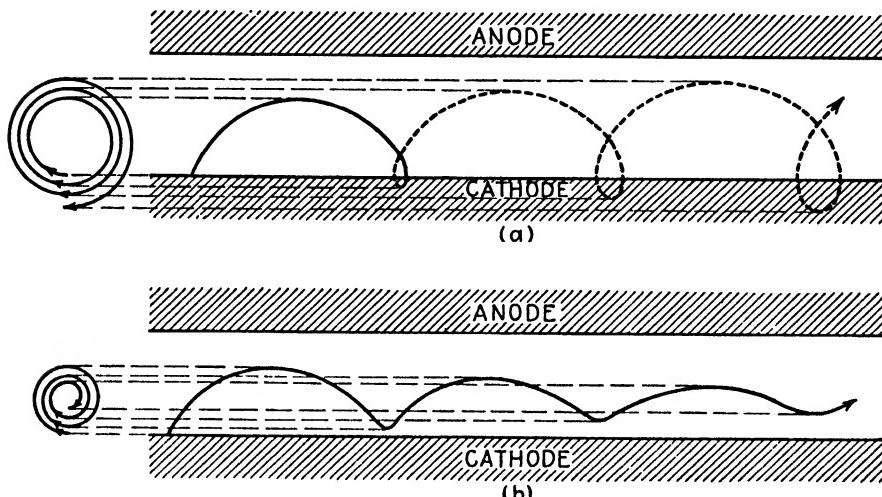
FIG. 12.—(a) Path of an electron which gains energy from the a-c field, and (b) path of an electron which gives energy to the a-c field. The spirals represent projections of the electron path.

whereby the unfavorable electrons (those which take energy from the a-c field) are withdrawn from the interaction space, while the favorable electrons (those which give energy to the a-c field) are allowed to continue to move through the interaction space, is similar to that described in Sec. 7.01. Briefly, this process may be stated as follows: Those electrons which take energy from the a-c field have their orbits altered in such a manner that they return to the cathode and thereby are withdrawn. The electrons which give energy to the a-c field have their orbits altered in such a manner that they remain in the interaction space. These two cases are illustrated for the parallel-plane magnetron in Figs. 12a and 12b, respectively.

Consider the electron paths in the parallel-plane magnetron under oscillating conditions. Equation (7.03-4) shows that the radius of curvature of an electron in a pure magnetic field varies directly with the velocity of the electron. Similarly, in a parallel-plane or cylindrical magnetron, the

radius of curvature varies more or less directly with the velocity of the electron. An electron which takes energy from the a-c field experiences an increase in velocity; hence the radius of curvature of the path increases as shown in the spiral of Fig. 12a.

Conversely, an electron which gives energy to the a-c field experiences a decrease in the radius of curvature of the electron path as shown in the spiral of Fig. 12b. The electron which gained energy from the a-c field would strike the cathode during the first cycle of oscillation and would thereby be removed, whereas the electron which gives energy to the a-c field would continue to travel through the interaction space. The cylindrical-anode magnetron can be vuisalized as a parallel-plane magnetron rolled into a circle, allowing the cathode radius to decrease. The electrons having unfavorable phase terminate at the cathode, while those oscillating in a favorable phase terminate at the anode.

In the cyclotron-frequency type of oscillation, the electrons in the interaction space gradually fall out of phase with the a-c field and it is necessary to provide some means of removing them before they begin to take energy from the a-c field. This can be accomplished by tilting the magnetron with respect to the magnetic field. The electrons are then given an axial component of motion, causing them to spiral down the interaction space and out the end of the tube. Another method consists of placing end plates perpendicular to the cathode and insulated from the cathode at either end of the tube. The end plates are given a positive potential so as to attract the electrons, thereby withdrawing them from the interaction space after several cycles of oscillation. The difficulty of removing the electrons at the appropriate time in their orbits, before they start taking energy from the a-c field, constitutes the principal drawback of the cyclotron-frequency oscillator. The efficiency of this type of oscillation is considerably lower than that for the traveling-wave type, being of the order of 10 to 15 per cent.

7.08. Traveling-wave Oscillation.[1-3]—Most of the magnetrons in present-day use are designed for operation in the traveling-wave modes of oscillation. To produce this type of oscillation, a multicavity magnetron such as that shown in Fig. 1 is required. The traveling-wave type of oscillation was briefly described in Sec. 7.01. As previously explained, the phase difference between the electrical oscillations of successive resonators is such as to produce a rotating a-c field or a traveling wave in the interaction

[1] Fisk, J. B., H. D. Hagstrum, and P. L. Hartman, The Magnetron as a Generator of Centimeter Waves, *Bell System Tech. J.*, vol. 25, pp. 167–348; April, 1946.

[2] Brillouin, L., Theory of the Magnetron, *Phys. Rev.*, part I, vol. 60, pp. 385–396; September, 1941; part II, vol. 62, pp. 166–167; August, 1942; part III, vol. 63, pp. 127–136; February, 1943.

[3] Brillouin, L., Practical Results from Theoretical Studies of Magnetrons, *Proc I.R.E.*, vol. 32, pp. 216–230; April, 1944.

space.   The electron space-charge cloud whirls around in the interaction space with a mean angular velocity equal to the angular velocity of a component of the rotating field.   Those electrons which rotate in such a phase as to take energy from the a-c field have their paths altered in such a manner that they spiral back to the cathode and are therefore withdrawn from the interaction space.   The remaining electrons are grouped in a spoke-like formation, with the spokes rotating synchronously with the a-c field and in such a phase that the electrons are retarded by the tangential component of the a-c field.   The paths of the individual electrons are approximately epicycloids, and the electrons which have a favorable phase eventually terminate at the anode.

Since the electrons are retarded by the tangential component of the a-c field, it would appear that they would slip behind the rotating field 'and gradually fall into an unfavorable phase in a manner similar to that of the cyclotron-frequency oscillation.   However, in the traveling-wave type of oscillation there is a "phase-focusing" effect due to the radial component of the a-c field which tends to keep the electrons rotating synchronously with respect to the rotating a-c field.   If an electron "leads" the retarding a-c field, the force due to the radial component of the a-c field is directed inward (toward the cathode).   The resulting radial component of velocity in the magnetic field produces a torque in such a direction as to decrease the angular velocity of the electron.   Conversely, an electron that "lags" the retarding a-c field will experience a force that is directed radially outward because of the radial component of the a-c field, and a torque, caused by the magnetic field, which tends to increase the angular velocity.   Thus, the phase-focusing action tends to retard the leading electrons and speed up the lagging electrons, thereby keeping the space-charge cloud rotating in synchronism with the retarding a-c field.

In the cyclotron type of oscillation, the a-c field is radial, whereas in the traveling-wave type of oscillation, the a-c field is largely tangential.   In traveling-wave modes, the electron paths would be similar to those shown in Fig. 12, except that the path of the unfavorable electron, Fig. 12a, would be tilted downward, whereas the path of the favorable electron, Fig. 12b, would be tilted upward.   The favorable electrons therefore eventually terminate at the anode and it is not necessary to make special provision to remove the electrons from the interaction space at a certain point in the electron cycle, as was required for the cyclotron-frequency oscillation.

The operation of the magnetron as a traveling-wave oscillator resembles, in many respects, the operation of a polyphase a-c generator.   In this analogy, the resonators of the magnetron correspond to the armature poles on the stator of the generator.   The space-charge spokes in the magnetron are analogous to a salient-pole rotor in the generator.   In the case of the

generator, the rotating magnetic field set up by the rotor induces emfs in the armature windings on the stator. The polyphase currents flowing in the stator windings produce a rotating magnetic field in the air gap between the rotor and stator. The reaction of this stator field back upon the field of the rotor tends to retard the rotor. If the rotor is to continue to revolve at constant angular velocity, it is necessary to supply sufficient mechanical power to the rotor to overcome the retarding force due to the armature field, as well as to make up for the mechanical losses in the generator.

Considering now the magnetron, we find that the rotating space-charge bunches induce emfs (and currents) in the resonators, analogous to the emfs induced in the stator windings of the generator by the rotor field. The resulting electrical oscillation of the resonators sets up a rotating electric field in the interaction space which reacts back upon the electrons, tending to slow them down. This is analogous to the reaction of the stator field back upon the rotor of the generator. In the magnetron the electron bunches tend to lead the retarding a-c field, while the phase-focusing effect, previously described, tends to keep the electrons rotating at synchronous speed with respect to the rotating field. The electrons gain energy from the d-c field as they spiral outward. This is comparable to the mechanical power supplied to the shaft of the generator.

. **Analysis of Traveling-wave Modes of Oscillation.**[1]—In the traveling-wave type of oscillation of a magnetron, the electrons rotate at synchronous speed with respect to a component of the a-c field. Our analysis will therefore deal, on the one hand, with the evaluation of the angular velocities of the electrons in terms of the magnetic field strength, d-c potential, etc., and, on the other hand, with the resonant frequencies of resonator systems and the traveling waves which are set up by the resonator fields.

Consider, for example, the magnetron shown in the developed view in Fig. 13. Each curve represents a plot of the a-c component of potential difference between the cathode and a point on the anode circle. The potential curves are plotted as a function of angle, with the various curves representing potential distributions at different instants of time. The potential is constant across the face of the anode and varies linearly across the gap.

The traveling-wave type of oscillation may have a number of different *modes* of oscillation. These correspond to the various resonant frequencies of the resonator system as well as to various angular velocities of the electrons which will react favorably with the field to produce oscillation. The particular mode shown in Fig. 13 is known as the π mode. This mode is

─────────

[1] The treatment of the traveling-wave modes of oscillation presented here is similar to that given in reference 1, *loc. cit.*

characterized by a phase difference of π radians between the electrical oscillation of successive resonators. The electric intensity (and hence the force on the electron) are proportional to the slope of the potential curve. A negative slope of the potential curve indicates that an electron in that particular position is being retarded by the a-c field.

The dotted lines in Fig. 13 represent the progress of electrons which travel in such a manner as to be retarded by the tangential component of
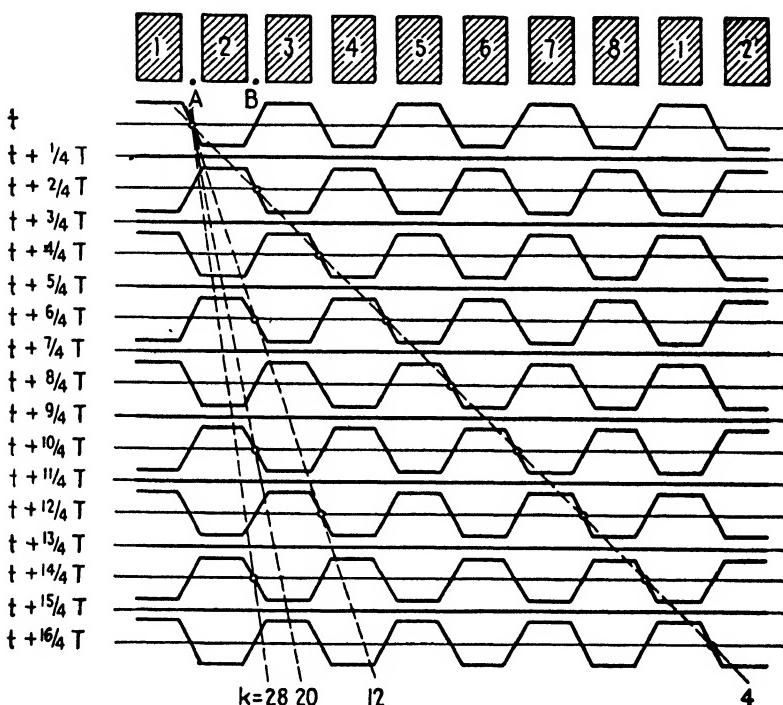
FIG. 13.—Potential distribution in an eight-resonator magnetron oscillating in the π mode. The various curves represent different instants of time. Dotted lines represent the progress of electrons that travel in a favorable phase.

the a-c field of successive resonators. These electrons, therefore, give energy to the a-c field, thereby contributing to the power output of the magnetron. The angular velocity of a particular electron is inversely proportional to the slope of the dotted line in Fig. 13. An electron may have any one of a number of discrete angular velocities (corresponding to the various dotted lines in Fig. 13) and still be retarded by the a-c field of successive resonators.

A convenient relationship may be derived, relating the electrical frequency of oscillation to the angular velocity of the electron. An electron, having an angular velocity $d\phi/dt$, travels once around the interaction space

(2π radians) in a time $2\pi/(d\phi/dt)$ seconds. For a magnetron having N anodes, the time T_1 required for the electron to traverse the distance from the center of one anode gap to the center of the next adjacent anode gap (from A to B in Fig. 13) is

$$T_1 = \frac{2\pi}{(d\phi/dt)N} \tag{1}$$

Now consider the phase relationships of the a-c field. Since the field at any point in the interaction space is single valued, the total phase shift around a closed path is $2\pi n$ radians, where n is an integer representing the number of cycles of phase shift around the closed path. Since there are N resonators, the phase difference between the oscillations of two adjacent resonators is

$$\theta = \frac{2\pi n}{N} \tag{2}$$

where n may take integer values from zero to $N/2$.

Assume that the electron has the same direction of rotation as the a-c field. The traveling electron then sees a phase difference between corresponding points of two successive resonators (points A and B in Fig. 13) of $\omega T_1 - \theta$ radians. Since the electron is to experience a maximum retarding force in both positions, this phase shift must be an integral multiple of 2π radians, or

$$\omega T_1 - \theta = 2\pi p \qquad p = 0, \pm 1, \pm 2, \cdots \tag{3}$$

By substituting θ from Eq. (2) and T_1 from Eq. (1) into Eq. (3) and solving for ω, we obtain for the angular frequency of electrical oscillation,

$$\omega = \left| \left(p + \frac{n}{N} \right) N \right| \frac{d\phi}{dt} = |k| \frac{d\phi}{dt} \tag{4}$$

where

$$k = \left(p + \frac{n}{N} \right) N \tag{5}$$

Equation (4) expresses a relationship between the oscillating frequency of the magnetron and the angular velocity of the electron. The constant k may take any one of a number of values as given by Eq. (5), the various values representing different modes of oscillation.

The value of ω in Eq. (4) must also coincide with a resonant frequency of the resonator system. A system of coupled resonators, such as that used in a magnetron, has a number of different resonant frequencies; hence, ω may represent any one of the resonant angular frequencies.

Under d-c operating conditions, the angular velocity of the electrons, as expressed by Eq. (7.05-3), was found to increase from a value of zero at

the cathode to a limiting value of $d\phi/dt = eB/2m$ at distances where $\rho \gg a$. Under dynamic conditions, however, most of the electrons rotate with constant angular velocity, in synchronism with the a-c field. Since the favorable electrons are retarded by the a-c field, it is apparent that the synchronous angular velocity must be less than the limiting value for d-c operating conditions.

An approximate expression for $d\phi/dt$ may be obtained by assuming that, at some radius ρ', the radial forces on an electron due to the electric and magnetic fields are equal and opposite. The force on the electron due to the d-c field, neglecting the variation of the field with radius, is $f = -eE = eV_0/(b - a)$. The radial force due to the motion of the electron in the magnetic field is $f = ev_\phi B = eB\rho' \, d\phi/dt$. Equating these forces, and solving for the angular velocity at radius ρ', we obtain

$$\frac{d\phi}{dt} = \frac{V_0}{\rho'B(b - a)} \tag{6}$$

Inserting this angular velocity into Eq. (4), we may solve for the value of V_0/B required for a given frequency of oscillation

$$\frac{V_0}{B} = \frac{\omega\rho'(b - a)}{|k|} \tag{7}$$

where ρ' is the as yet undetermined radius at which the radial forces on the electron are equal and opposite. As a first approximation, we may assume that this occurs at the mid-point between cathode and anode; *i.e.*, at $\rho' = (a + b)/2$, giving

$$\frac{V_0}{B} = \frac{\omega(b^2 - a^2)}{2|k|} \tag{8}$$

Equation (8) may be used to determine the approximate frequency of oscillation if V_0 and B are known.

Hartree has given a more accurate derivation, similar to Eq. (8), based upon the assumption that the electrons just reach the anode for an infinitesimal amplitude of a-c voltage. His equation is

$$V_0 = \frac{\omega b^2 B}{2|k|}\left[1 - \left(\frac{a}{b}\right)^2\right] - \frac{m}{2e}\left(\frac{\omega b}{|k|}\right)^2 \tag{9}$$

This is a quadratic equation in terms of frequency, which may be used to evaluate the frequency if the values of V_0 and B are known. If the final term in Eq. (9) is neglected, this equation reduces to Eq. (8).

From an operational viewpoint, it is necessary to adjust the magnetic field strength and d-c anode potential of the magnetron to obtain the proper value of $d\phi/dt$ which will satisfy Eq. (4) for the desired mode.

There is a large number of possible modes of oscillation, as expressed by Eq. (9), although only a few of the modes have any practical significance, since many of the modes yield low power output and low efficiency. Our principal concern with the extraneous modes is a defensive one, arising out of the fact that magnetron oscillations have a tendency to jump from one mode to another mode under slight changes in operating conditions. This results in an abrupt change in frequency of oscillation and power output. This tendency of "mode-jumping" can be remedied by methods which will be described later.

It should be noted that the value of k in Eqs. (4) and (5) may be either positive or negative, whereas ω is always taken as a positive quantity.

7.10. The π Mode.—The most common mode of oscillation is the π mode, for which the phase shift θ between successive resonators is π radians. From Eq. (7.09-2) we obtain, for the π mode, $n = N/2$. Inserting this into Eq. (7.09-5), the allowed values of k become

$$k = (p + \tfrac{1}{2})N \tag{1}$$

Usually the π mode is operated in such a manner that the electrons rotate synchronously with respect to the fundamental component of potential. This requires that $p = 0$ and hence $k = N/2 = n$. Inserting this value of k into Eq. (7.09-4), we obtain the angular velocity of the electrons $d\phi/dt = 2\omega/N$. For example, the angular velocity of an electron in an eight-resonator magnetron would be $d\phi/dt = \omega/4$ radians per second.

Let us now consider the a-c field of the π mode. As shown in Fig. 13, the potential distribution as a function of the angle is trapezoidal. If we choose our reference at the center of one of the anode pole faces, then the Fourier series representing the potential as a function of ϕ contains only cosine terms, thus

$$V_{ac} = \sum_{k=1}^{k=\infty} V_k \cos \omega t \cos k\phi \tag{2}$$

where V_k is the amplitude of the kth harmonic. In the Fourier series representing the a-c potential, k must have integer values from zero to infinity. However, we are interested only in those components in the Fourier series which react favorably with the electrons; hence we shall consider only those values of k which satisfy Eq. (1).

Equation (2) may be expanded by a trigonometric identity to give

$$V_{ac} = \sum_{k=1}^{k=\infty} \frac{V_k}{2} \left[\cos(\omega t - k\phi) + \cos(\omega t + k\phi) \right] \tag{3}$$

The first term in Eq. (3) represents a potential wave traveling in the $+\phi$ direction with an angular velocity ω/k radians per second. The second

term represents a wave traveling in the $-\phi$ direction, also with the angular velocity ω/k radians per second.

The potential distribution shown in Fig. 13 constitutes a *standing wave*. This is expressed as a Fourier series in Eq. (2). In Eq. (3), each one of the Fourier-series components representing a *standing wave* is resolved into two *traveling waves* which travel in opposite directions with equal angular velocities. From the point of view of magnetron operation, the electrons may rotate approximately synchronously with respect to any one of the traveling-wave components in either direction of travel.

In order for the electrons to react favorably with the fundamental component of the potential wave, we must have $p = 0$ and hence, from Eq. (1), $k = N/2 = n$. The angular velocity of the electrons is then $d\phi/dt = \omega/k = 2\omega/N$. The electrons may rotate in either direction and thus travel approximately synchronously with respect to either one of the traveling waves corresponding to the fundamental component. The electrons will react with a harmonic of the potential wave when they travel approximately synchronously with respect to the traveling wave corresponding to that particular harmonic. The harmonics of the potential wave with which the electrons can react favorably are known as the *Hartree harmonics*.

On the average, an electron will transfer power only to that particular component of the field which has approximately the same angular velocity as the electron. The power transfer to all other field components will rapidly alternate between positive and negative values, averaging zero.

7.11. Other Modes of Oscillation.—The mode corresponding to $n = 0$ is the cyclotron-frequency mode which was previously discussed. For this mode, we have $\theta = 0$; hence, the potential of all of the anodes rises and falls in time phase and the a-c field is essentially radial. The value of k is $k = pN$ and Eq. (7.09-4) gives $\omega = pN(d\phi/dt)$.

For the π mode, we found that the potential distribution can be represented by a Fourier series. Any one harmonic component in the series may be regarded either as a standing wave or as two component traveling waves which travel in opposite directions with equal angular velocities.

In the more general case, the potential distribution can be represented by a similar Fourier series of component waves traveling in opposite directions, thus

$$V_{ac} = \sum_{k=1}^{k=\infty} [A_k \cos(\omega t - k\phi + \delta) + B_k \cos(\omega t + k\phi + \gamma)] \qquad (1)$$

where δ and γ are arbitrary phase constants. This expression differs from Eq. (7.10-3) for the π mode in that the two oppositely directed waves for the π mode have equal amplitudes, whereas in the general case, the amplitudes may be unequal.

In the case of the π mode, there is no preferred direction of rotation of the electrons. Either direction will give the same power output. However, in the more general case, the preferred direction is the direction of rotation of the stronger potential wave. When the electrons travel in the direction of rotation of the weaker potential wave, they are referred to as driving a "reverse mode."
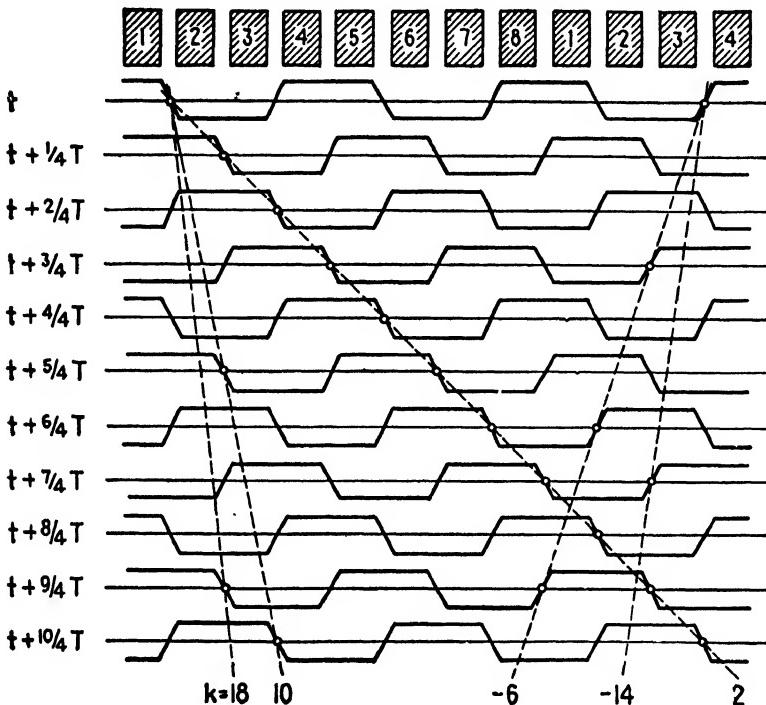
Fig. 14.—Anode potential as a function of angular position and time for the modes corresponding to $n = 2$. Dotted lines represent paths of favorable electrons. Negative values of k correspond to electrons which are driving a reverse mode.

Figure 14 shows a traveling wave of potential as a function of angular position and time for a magnetron having eight resonators. The dotted lines represent favorable electron paths corresponding to various modes. The electrons corresponding to the $-k$ modes are driving a reverse mode.

For the π mode, the positive and negative values of p in Eq. (7.09-5) give the same sequence of values of k. However, for other modes, the sequence of values of k are different for positive and negative values of p. For example, if we let $N = 8$ and $n = 2$, the positive values of p give $k = 2, 10, 18 \cdots$ while the negative values of p give $k = -6, -14, -22$.

The electric and magnetic field distributions for various modes in a magnetron having eight resonators are shown in Fig. 15. The electric field
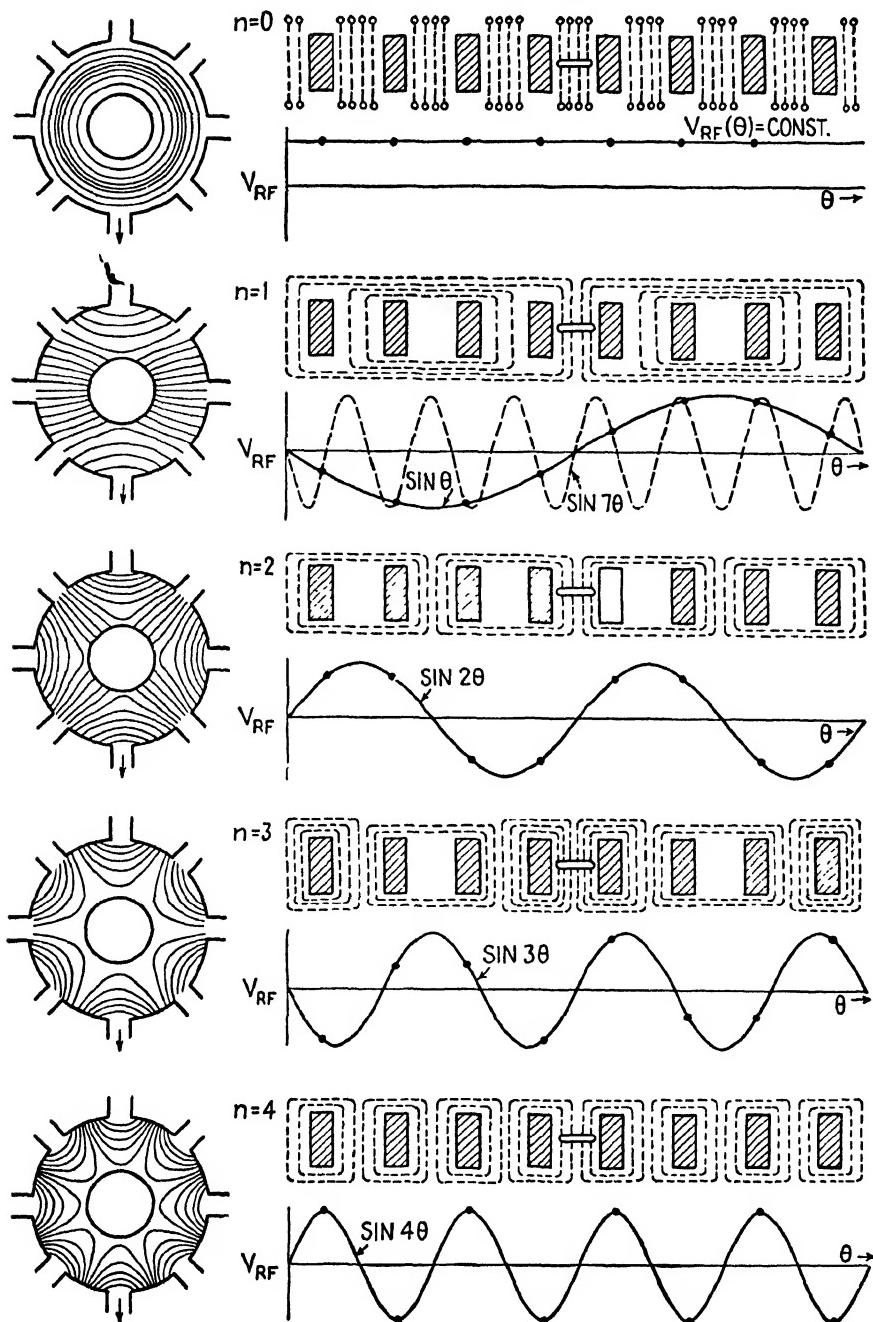
FIG. 15.—Electric and magnetic field distributions in an eight-resonator magnetron for various modes. The mode $n = 4$ is the π mode.

is represented by the solid lines in the circular diagram, whereas the magnetic field is represented by the dotted lines in the developed view. The output coupling loop is represented by the arrow in the circular drawing and by the bar in the center of the developed view. The sine-wave curves represent the fundamental component of potential as a function of θ. For the mode $n = 1$, the potential distribution for the second Hartree harmonic $(p = -1, k = -7)$ is also shown.

7.12. Resonant Frequencies of the Resonator System.—For the purpose of studying the resonant frequencies of the resonator system, let us represent a single resonator by a parallel L-C circuit. Since a mutual coupling exists between the various resonators in the magnetron, the equivalent circuit would consist of a closed chain of parallel L-C circuits, with mutual inductance between adjacent circuits.

If two identical parallel L-C circuits are coupled together and shock excited, the system will oscillate simultaneously at two slightly different frequencies. These two frequencies produce a beat effect, with part of the energy surging back and forth between the two coupled circuits at the beat frequency. The two frequencies of oscillation result from the fact that the mutual inductance can either add to or subtract from the self-inductances. Hence, the two circuits may oscillate either in phase or π radians out of phase, the two cases corresponding to two slightly different resonant frequencies.

A closed chain of N identical resonators, such as that used in the magnetron, would have N resonant frequencies, spaced a short distance apart on the frequency scale. In the case of the two-resonator system, the oscillations of the two resonators differed by a phase angle of either 0 or π radians. In the more general case of N resonators arranged in a ring-shaped configuration, the phase difference between successive resonators is not restricted to 0 or π radians, but may have any value provided that the sum of the phase differences around the closed system shall equal some integral multiple of 2π radians. Instead of having the mutual inductance either add to or subtract from the self-inductance, in the more general case, there will be a phase difference between the effects of self and mutual inductance which makes new resonant frequencies possible. As the coupling between resonators decreases, the various resonant frequencies draw closer together, finally converging upon the resonant frequency of a single resonator.

7.13. Mode Separation.—In the design of a magnetron, it is necessary to provide some means of compelling the magnetron to oscillate in a single mode in order to avoid "mode jumping." To a certain extent, this can be accomplished by using tight coupling between resonators, thereby separating the resonant frequencies as far apart as possible on the frequency scale.

A method which is commonly used to minimize the possibility of mode jumping is to connect the anode segments together by means of conducting straps in such a manner as to maintain a fixed phase relationship between the oscillations of the various resonators. Figure 16 shows a magnetron with two ring straps. Each strap is connected electrically to alternate anodes so as to compel the potentials of the alternate anodes to oscillate in time phase. This restricts the possible modes of oscillation to the π mode or the mode $n = 0$ (for which the potentials of all anode segments oscillate in time phase). Each of the straps is broken at one point in order to allow for the asymmetry of the field produced by the coupling loop. The straps add a capacitative reactance and hence cause a shift in the mode-frequency distribution.
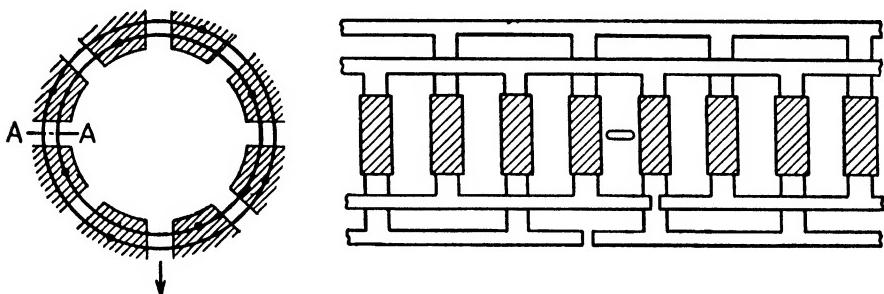
FIG. 16.—Strap arrangement for the π mode.

It has been found advisable to shield the straps from the interaction space in order to minimize the possibility of the $n = 0$ mode, since the strap situated closest to the cathode tends to set up a radial a-c field between cathode and anode which gives rise to the $n = 0$ mode. Shielding can be accomplished by milling grooves in the anode structure and embedding the straps in the grooves.

The "rising-sun" resonator system shown in Fig. 17 provides a means of obtaining mode separation without the use of straps. The resonator system contains alternately large- and small-size resonators. The two sizes of resonators, taken individually, have widely differing resonant frequencies. When the resonators are coupled together, as in the magnetron, the π mode resonant frequency lies between the two individual resonant frequencies. There are also a number of other resonant modes present in the coupled system. However, if the two sizes of resonators differ appreciably, there will be sufficient mode separation to assure relatively stable operation in the π mode.

Figure 18 shows a comparison of the mode separation of: (a) an unstrapped magnetron with a conventional resonator system, (b) a heavily strapped magnetron, and (c) a magnetron containing a rising-sun resonator
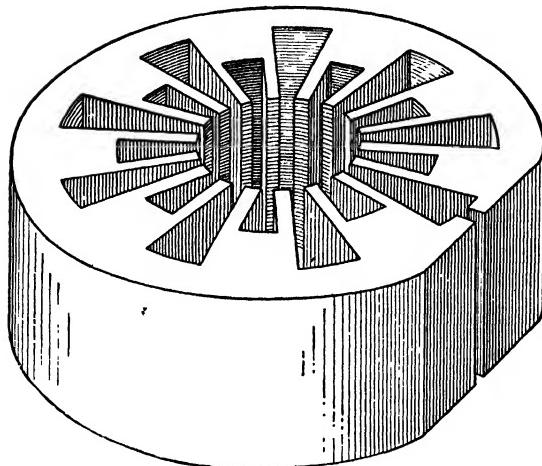
Fig. 17.—The "rising-sun" resonator.
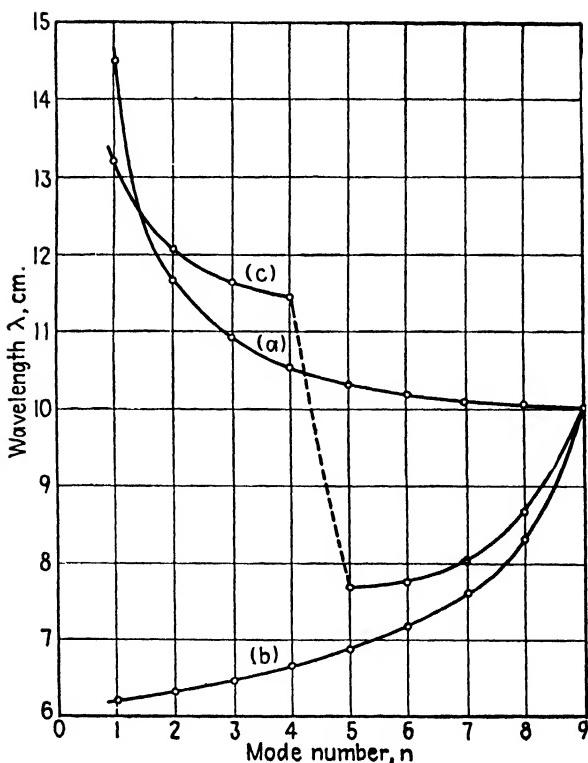
Fig. 18.—Plot of wavelengths as a function of mode number for magnetrons having 18 resonators. (a) Unstrapped resonator system, (b) heavily strapped resonator, and (c) "rising-sun" resonator. The π mode corresponds to $n = 9$. (*Courtesy of the Bell System Tech. Jour.*)

system. The resonant frequencies of the rising-sun resonator fall into two groups, with the π mode approximately halfway between the two groups. The mode separation for this particular rising-sun resonator system is not quite as great as that of the heavily strapped magnetron, but there are compensating advantages in favor of the rising-sun resonator.
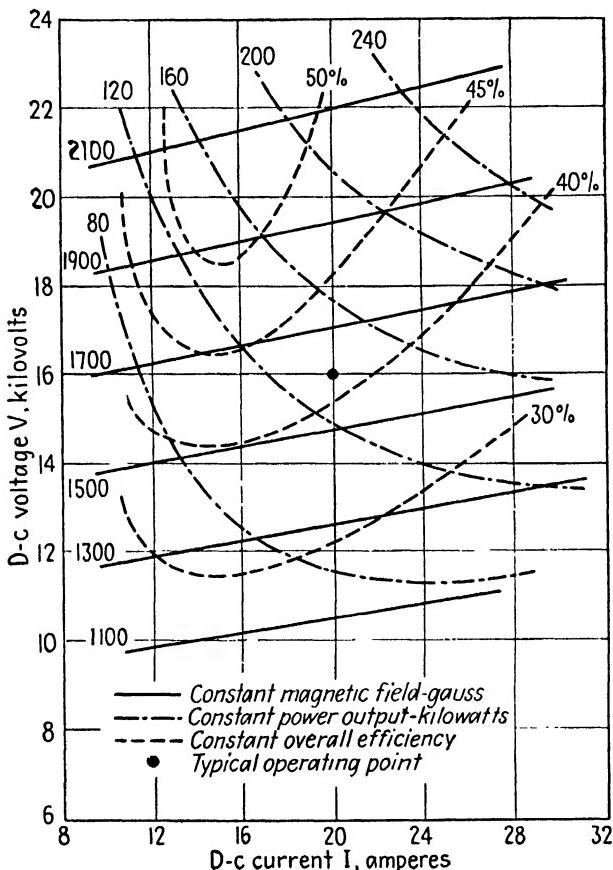
Fig. 19.—Contours of constant magnetic field strength, constant power output, and constant efficiency, plotted against voltage and current for a 10-centimeter pulsed magnetron. (*Courtesy of the Bell System Tech. Jour.*)

In the strapped magnetron, the mode separation decreases as the axial length of the anode increases. This is particularly objectionable in short-wavelength magnetrons, since the restriction on the anode length imposes a serious limitation upon the power output obtainable from the magnetron. In the rising-sun magnetron, the mode separation is not seriously impaired by anode lengths up to approximately $\frac{3}{4}$ of a wavelength, which is somewhat greater than the allowable anode length for the strapped magnetron. Furthermore, in the rising-sun magnetron, a large number of resonators may

be used and still maintain a reasonable degree of mode separation, whereas in the strapped magnetron, the modes tend to fall close together when a large number of resonators are used.   The copper losses in the rising-sun magnetron are somewhat less than those in the strapped magnetron; hence the efficiency of the rising-run magnetron is higher.

The principal disadvantage of the rising-sun magnetron is the tendency to operate in the zero mode ($n = 0$).   Because of the asymmetry of the resonators, the a-c field strength across the gap of the large resonators
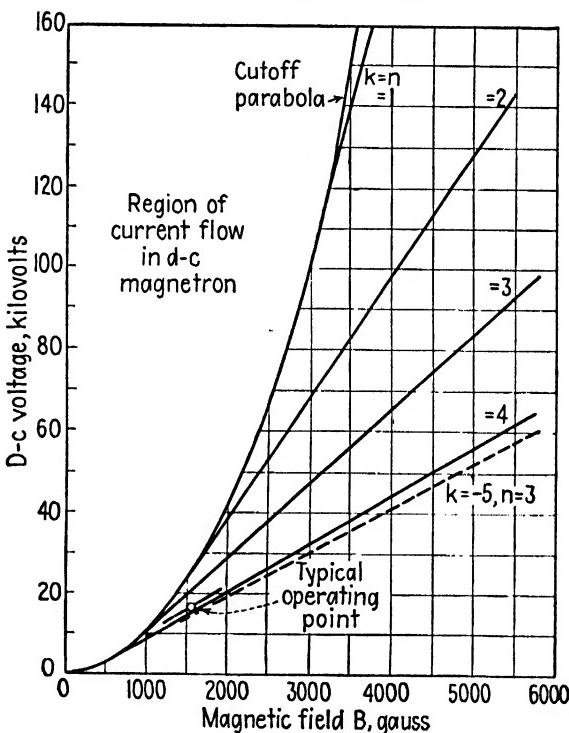
Fig. 20.—D-c anode potential against magnetic field strength for various modes of oscillation

differs from that of the small resonators.   Since all of the resonators of one size oscillate in time phase, the excess field contributes to a zero-mode field.   The magnetron can be designed so as to minimize the possibility of zero-mode oscillation by avoiding the value $\lambda B = 15,000$ (λ in centimeters, B in Gauss) required for zero-mode oscillation.

7.14. Representation of Performance Characteristics of Magnetrons.— The performance characteristics of a magnetron can be represented by a contour plot such as that shown in Fig. 19.   This graph shows contours of constant-power output, constant efficiency, and constant magnetic field strength as a function of d-c anode voltage and anode current.   The partic-

ular plot shown in Fig. 19 was obtained for an eight-resonator, 10-centimeter magnetron which was pulsed with a pulse of one microsecond duration and 1,000 pulses per second. The typical operating point represented by the black dot in the center of the graph corresponds to a peak power output of 135 kilowatts at 42 per cent efficiency, requiring an anode potential of 16 kilovolts.

Figure 20 shows a plot of d-c anode potential against magnetic field strength for various modes in an eight-resonator magnetron. The cutoff
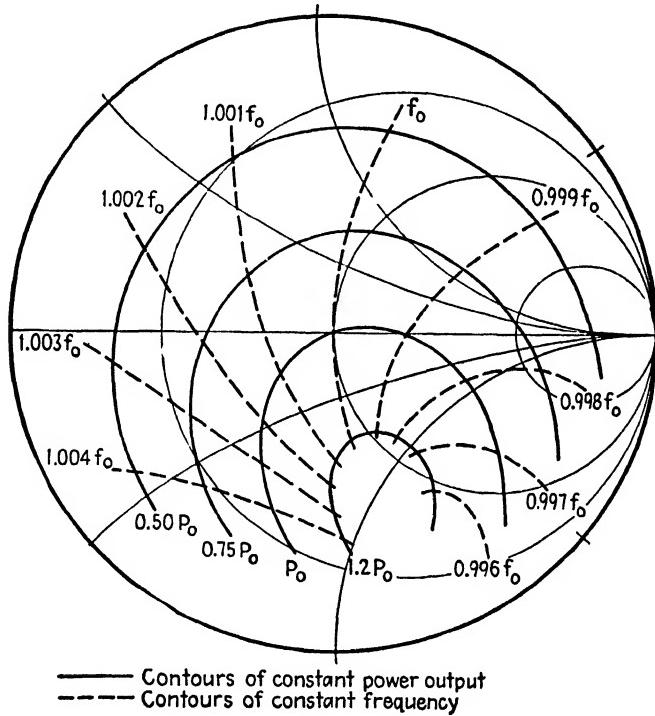
FIG. 21.—Rieke diagram showing contours of constant frequency and constant power output plotted on an admittance diagram similar to that in Fig. 3 [Chap. 9].

parabola is a plot of Eq. (7.05-10), whereas the straight lines representing the various modes are a plot of the Hartree equation (7.09-9). The magnetron is the same as that for which the performance characteristics are shown in Fig. 19. The magnetic field strengths required for the various modes of operation are somewhat greater than the cutoff value, as indicated by the typical operating point in Fig. 20. The range of values of V_0 and B in Fig. 20 is considerably greater than those used in ordinary practice.

The Rieke diagram is a method of representing the performance characteristics of a magnetron in terms of the load impedance (or admittance). Figure 21 shows a Rieke diagram consisting of contours of constant-power

output and constant frequency plotted on an admittance diagram similar to the Smith diagram described in Chap. 9. These data are obtained experimentally by varying the load impedance and observing the power output and frequency. The Rieke diagram shows how the operation of the magnetron is affected by load impedance, thereby making it possible to select the optimum load impedance.

7.15. Equivalent Circuit of the Magnetron.—A useful criterion of oscillation was stated in Chap. 5. Another useful criterion of oscillation is the admittance criterion, which states the requirements of oscillation in terms
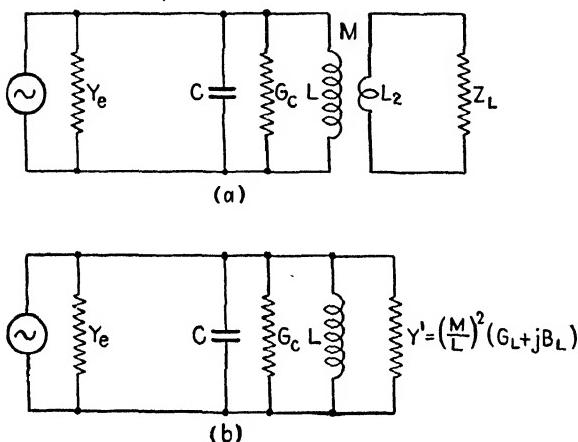
Fig. 22.—Equivalent circuit of the magnetron.

of the circuit parameters. Briefly, the admittance criterion of oscillation states that oscillation will occur if the sum of the admittances looking both ways at any pair of terminals is zero.

If we apply this criterion to the magnetron, we may take the junction at the extremities of the anode gap. One of the admittances is then the circuit admittance looking into the anode at the junction. The other admittance is the admittance of the electron stream. Representing these by Y_c and Y_e, respectively, the criterion of oscillation requires that

$$Y_c + Y_e = 0 \qquad (1)$$

For convenience, we represent the electron admittance by

$$Y_e = G_e + jB_e \qquad (2)$$

In order to facilitate the circuit analysis, the magnetron will be represented by the equivalent circuit shown in Fig. 22. In this equivalent circuit, L and C represent the combined inductance and capacitance of N resonators in parallel, as obtained by multiplying the inductance and

capacitance of a single resonator by $1/N$ and N, respectively. The shunt conductance G_c represents the loss in the resonator walls. The inductance of the coupling loop is represented by L_2 and the mutual inductance between the resonator and coupling loop is represented by M. The external load impedance is Z_L.

For an ideal transformer, the secondary admittances may be transferred to the primary by multiplying them by the factor $(M/L)^2$, yielding

$$Y' = \left(\frac{M}{L}\right)^2 \left(\frac{1}{jX_2 + Z_L}\right) = \left(\frac{M}{L}\right)^2 (G'_L + jB'_L) \tag{3}$$

where $X_2 = \omega L_2$, and G'_L and B'_L are, respectively, the conductance and susceptance of the secondary circuit.

The circuit admittance Y_c may be viewed as the ratio of the induced a-c current at the anode junction to the a-c voltage developed across the anode gap. This may be expressed in terms of the primary circuit parameters, shown in Fig. 22, and the reflected secondary circuit parameters as follows:

$$Y_c = \frac{1}{j\omega L} + j\omega C + G_c + Y'_L \tag{4}$$

Now let $\omega_0 = 1/\sqrt{LC}$ and $Y_0 = \sqrt{C/L}$, the latter being referred to as the characteristic admittance of the resonator. Substitution of these into Eq. (4) yields

$$Y_c = j\sqrt{\frac{C}{L}}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) + G_c + Y'_L$$

$$\cong 2jY_0\left(\frac{\omega - \omega_0}{\omega_0}\right) + G_c + Y'_L \tag{5}$$

Upon inserting Y'_L from Eq. (3) into (5), we obtain

$$Y_c = 2jY_0\left(\frac{\omega - \omega_0}{\omega_0}\right) + G_c + \left(\frac{M}{L}\right)^2 (G'_L + jB'_L) \tag{6}$$

Oscillation will occur if the sum of the circuit admittance and the electron admittance is zero. Adding Eqs. (2) and (6) and setting $Y_c + Y_e = 0$, we obtain the criterion for oscillation:

$$G_e = -G_c - \left(\frac{M}{L}\right)^2 G'_L \tag{7}$$

$$B_e = -2Y_0\left(\frac{\omega - \omega_0}{\omega_0}\right) - \left(\frac{M}{L}\right)^2 B'_L \tag{8}$$

The first of these states that the equivalent conductance of the electron stream must be equal in magnitude but of opposite sign to that of the circuit conductance. Since the circuit conductance is always positive, the electronic conductance must be negative in order for oscillations to occur. The second equation above states that the electronic and circuit susceptances must be equal in magnitude but opposite in sign. The frequency of oscillation is determined largely by this equation.

The foregoing relationships help one to visualize what adjustments are likely to take place under conditions of variable load impedance. For example, if the load susceptance B'_L is varied, it is possible for the angular frequency of oscillation ω to change in such a manner as to satisfy Eq. (8)
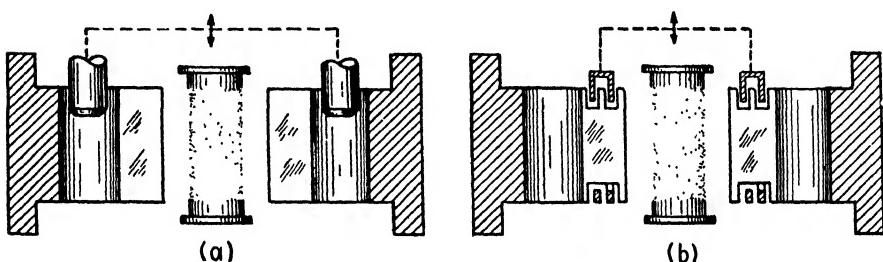
FIG. 23.—Tuning of magnetrons (a) by variable inductive plunger, and (b) by variable capacitive straps.

without appreciably altering the value of B_e. This accounts for the frequency drift of the magnetron with variable loading. It should be noted, however, that the electron admittance $Y_e = G_e + jB_e$, in general, is not constant, but rather varies with the operating conditions of the magnetron.

Variations in load impedance, particularly those due to variations in load reactance, tend to cause a change in the operating frequency of the magnetron. The frequency may be stabilized by using a high-Q resonator system or by increasing the characteristic admittance, $Y_0 = \sqrt{C/L}$, of the resonator. A high Q requires light loading and, hence, low power output. The over-all Q may be improved somewhat by coupling the output of the magnetron into a high-Q external resonator and then coupling this to the load.

7.16. Tunable Magnetrons.[1]—One of the early limitations of the magnetron was the fact that it constituted a fixed-frequency oscillator. Later, however, magnetrons were developed which could be tuned over a range of ± 7 to 20 per cent of the mean frequency.

The frequency of the magnetron may be varied either by varying the effective inductance or the effective capacitance of the resonators. Figure

[1] NELSON, R. B., Methods of Tuning Multiple-Cavity Magnetrons, *Phys Rev.*, vol. 70, p. 118; July, 1946

23 shows a means of varying the inductance by sliding a conducting pin into or out of the resonator. The effective capacitance can be varied by moving an annular ring, shaped like a cooky cutter, into or out of grooves in the anode. Both methods have been used successfully to obtain variable-frequency magnetrons. A combination of both the inductive tuning and capacitive tuning may be used to obtain a still wider range of frequencies. The frequency of a magnetron may also be varied over a narrow range by varying the anode potential or by coupling a tunable resonator to the output and tuning this to a slightly different resonant frequency.

PROBLEMS

1. Referring to Eqs. (7.04-11 and 12), let $k = eV_0/\omega_c^2 md$. Plot curves of z/k and x/k against $\omega_c t$. These curves can be used to determine the position of an electron as a function of time in a parallel-plane magnetron with any value of applied voltage and flux density.

2. A split-anode magnetron has dimensions $a = 0.02$ cm, $b = 0.3$ cm, $l = 1$ cm. The magnetic flux density is 3,000 gauss. Compute the following:

(a) Wavelength for the cyclotron mode of oscillation.

(b) Approximate d-c voltage required.

(c) Approximate angular velocity of the electrons.

3. A magnetron, operating in the π mode ($p = 0$), has the following characteristics:

$$N = 8 \qquad\qquad f = 2,800 \text{ mc}$$
$$a = 0.3 \text{ cm} \qquad\qquad V_0 = 16 \text{ kv}$$
$$b = 0.8 \text{ cm} \qquad\qquad B = 0.16 \text{ webers}/m^2$$
$$l = 2.0 \text{ cm (anode length)}$$

(a) Using Eq. (7.09-4), compute the angular velocity of the electrons and compare this value with that obtained by using Eq. (7.05-3).

(b) Determine the radius ρ' in Eq. (7.09-7) at which the radial forces due to the electric and magnetic fields are equal and opposite.

(c) Assume that the radius ρ' computed in part (b) is valid for all higher modes. Using Eq. (7.09-7), determine the values of V_0/B required for a frequency of 2,800 mc in each of the following modes:

(1) $n = 0$, $p = \pm 1$ (the cyclotron mode)

(2) $n = 1$, $p = 0$, ± 1

(3) $n = 2$, $p = 0$, ± 1

Explain the physical interpretation of each mode.

4. An atomic model can be constructed by reversing the potentials on the electrodes in a magnetron; i.e., by using a central anode and a concentric cathode. If electrons are projected into the interelectrode space so as to miss the anode, they will rotate around the anode much as electrons rotate about the nucleus in an atom. Write the equations of motion of the electrons and discuss the possibilities of using this principle for an oscillator.

CHAPTER 8

TRANSMISSION-LINE EQUATIONS

The transmission line may be analyzed by the solution of Maxwell's field equations or by the methods of ordinary circuit analysis. The solution of the field equations involves the determination of the field intensities in three-dimensional space; hence three space variables in addition to the time variable are involved. Although this method may be used to analyze a few systems having relatively simple geometry, in most practical cases the mathematical complications resulting from four independent variables are usually insurmountable. The solution of Maxwell's field equations reveals that the energy propagates through the dielectric medium as an electromagnetic wave, the conductors serving to guide the energy flow.

In the circuit method, the effects of the electric and magnetic fields are taken into consideration by the use of the circuit parameters, *i.e.*, the capacitance, inductance, resistance, and conductance. By this procedure, the mathematical analysis is reduced to a problem involving one space variable in addition to the time variable. The circuit method, however, does not yield the complete solution. In a later chapter we shall find, using the Maxwellian method, that an infinite number of electromagnetic field configurations, known as modes, may be associated with a given transmission line. The *principal mode* corresponds to the field configuration which exists at frequencies for which the spacing between conductors is appreciably less than a quarter wavelength. The *higher modes* appear when the separation distance between conductors is of the order of magnitude of a quarter wavelength or greater, or when there are impedance discontinuities on the line.

In this chapter we shall analyze the transmission line using the circuit method. The Maxwellian method will be considered in later chapters.

8.01. Derivation of the Transmission-line Equations.—Consider the transmission line of Fig. 1, which is assumed to have uniformly distributed parameters. The resistance, inductance, capacitance, and conductance per unit length are represented by R, L, C, and G, respectively.

The equations for the instantaneous voltage Δv across an incremental length of line Δx, and the shunt current through it are

$$\Delta v = i(R\,\Delta x) + (L\,\Delta x)\frac{\partial i}{\partial t} \tag{1}$$

$$\Delta i = v(G\,\Delta x) + (C\,\Delta x)\frac{\partial v}{\partial t} \tag{2}$$

146

Replacing Δv in Eq. (1) by $(\partial v/\partial x)\,\Delta x$ and dividing by Δx, with a similar operation for the current equation, we obtain the differential equations of the transmission line,

$$\frac{\partial v}{\partial x} = iR + L\frac{\partial i}{\partial t} \tag{3}$$

$$\frac{\partial i}{\partial x} = vG + C\frac{\partial v}{\partial t} \tag{4}$$
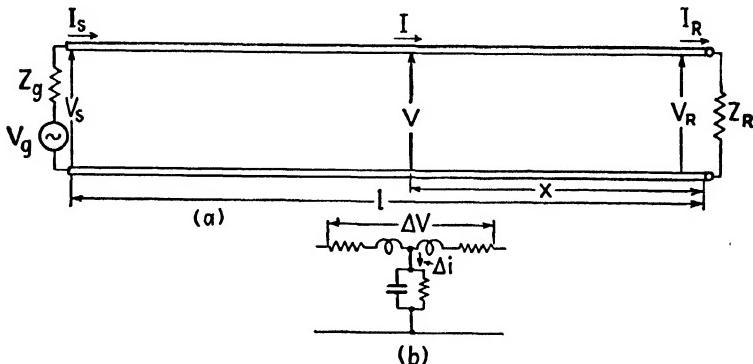
FIG. 1.—(a) Transmission line, and (b) equivalent circuit of a differential element of line.

In order to illustrate the nature of the solution, consider the case of a lossless line in which we have $R = G = 0$. Equations (3) and (4) then reduce to

$$\frac{\partial v}{\partial x} = L\frac{\partial i}{\partial t} \tag{5}$$

$$\frac{\partial i}{\partial x} = C\frac{\partial v}{\partial t} \tag{6}$$

Now differentiate Eq. (5) with respect to x and substitute Eq. (6) for $\partial i/\partial x$ in the resulting equation. Similarly, differentiate Eq. (6) with respect to x and substitute Eq. (5) for $\partial v/\partial x$. This process gives

$$\frac{\partial^2 v}{\partial x^2} = LC\frac{\partial^2 v}{\partial t^2} \tag{7}$$

$$\frac{\partial^2 i}{\partial x^2} = LC\frac{\partial^2 i}{\partial t^2} \tag{8}$$

These are known as the *wave equations* for the lossless line. Equations of this type frequently occur in the analysis of electrical, mechanical, and acoustical systems. Solutions of Eqs. (7) and (8) are of the form $f_1[t - (x/v)]$ or $f_2[t + (x/v)]$.

The function $f_1[t - (x/v)]$ represents a wave of arbitrary waveform (determined by the particular function chosen) which travels in the $+x$ direction with a velocity v. Similarly the function $f_2[t + (x/v)]$ represents a wave traveling in the $-x$ direction with a velocity v. Figures 2a and 2b illustrate waves traveling in the $+x$ and $-x$ direction, respectively, for several successive instants of time. Consider the function $f_1[t - (x/v)]$. If we were to ride along with the peak of the wave, it would be necessary for our displacement x to vary with time in such a manner as to hold $t - (x/v)$ constant. Thus, as time t increases, x must increase in a positive direction. We therefore conclude that $f_1[t - (x/v)]$ represents a wave traveling in the $+x$ direction.

(a)-Wave traveling in +x direction representing $f(t-\frac{x}{v})$

(b)-Wave traveling in -x direction representing $f(t+\frac{x}{v})$
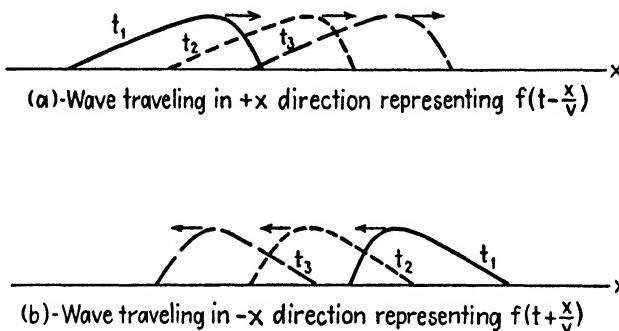
Fig. 2.—Traveling waves.

To find the velocity v, let us substitute $v = f_1[t - (x/v)]$ into Eq. (7). We obtain $\partial^2 v/\partial x^2 = (1/v^2)f_1''[t - (x/v)]$ and $\partial^2 v/\partial t^2 = f_1''[t - (x/v)]$. Inserting these into Eq. (7) and solving for the velocity, we obtain

$$v = \frac{1}{\sqrt{LC}} \tag{9}$$

For a lossless line with a dielectric having a relative permittivity of unity, the velocity v is equal to the velocity of light, *i.e.*, $v = 3 \times 10^8$ meters per second.

8.02. Sinusoidal Impressed Voltage.—In most practical applications, we are concerned with voltages and currents having a sinusoidal time variation. For mathematical convenience, such a variation may be represented by the time function $e^{j\omega t}$. We therefore let [1]

$$v = V e^{j\omega t} \tag{1}$$

$$i = I e^{j\omega t} \tag{2}$$

[1] An instantaneous voltage of the form $v = V_m \cos(\omega t + \theta)$ may be written $v = \text{Re} V_m e^{j(\omega t+\theta)}$, where Re signifies that we take the real part of the quantity following it. Thus, expanding $e^{j(\omega t+\theta)} = \cos(\omega t + \theta) + j \sin(\omega t + \theta)$ and discarding the imaginary part, we have $\text{Re} e^{j(\omega t+\theta)} = \cos(\omega t + \theta)$. Now write $v = \text{Re}[(V_m e^{j\theta})e^{j\omega t}]$

where V and I are complex quantities which are functions of x but not of time t. Inserting Eqs. (1) and (2) into (8.01-3 and 4), the time function $e^{j\omega t}$ cancels out and we obtain equations in which the voltage and current are functions of the space variable alone,

$$\frac{dV}{dx} = Iz \qquad (3)$$

$$\frac{dI}{dx} = Vy \qquad (4)$$

where

$$z = R + j\omega L \qquad (5)$$

$$y = G + j\omega C \qquad (6)$$

represent, respectively, the series impedance and shunt admittance per unit length of line.

To obtain an explicit equation for voltage, differentiate Eq. (3) with respect to x and substitute dI/dx from Eq. (4). The current equation is obtained by differentiating Eq. (4) with respect to x and substituting dV/dx from Eq. (3). These operations yield

$$\frac{d^2V}{dx^2} = \gamma^2 V \qquad (7)$$

$$\frac{d^2I}{dx^2} = \gamma^2 I \qquad (8)$$

where $\gamma = \sqrt{zy}$ is the *propagation constant* of the line. In general, γ is complex and may be separated into real and imaginary parts, hence we let

$$\gamma = \sqrt{zy} = \alpha + j\beta \qquad (9)$$

where α is known as the *attenuation constant* and β is the *phase constant*.

The solutions of Eqs. (7) and (8) which also satisfy (3) and (4) are

$$V = Ae^{\gamma x} + Be^{-\gamma x} \qquad (10)$$

$$I = \frac{A}{Z_0} e^{\gamma x} - \frac{B}{Z_0} e^{-\gamma x} \qquad (11)$$

where

$$Z_0 = \sqrt{\frac{z}{y}} \qquad (12)$$

is the *characteristic impedance* of the line.

---

where θ is the phase angle of the voltage at zero time. Letting $V = V_m e^{j\theta}$, we have $v = \text{Re} V e^{j\omega t}$. It is customary to drop the designation Re although it is implied and should be reinserted if we wish to obtain actual values of the voltage or current at any instant of time. Thus, we have $v = V e^{j\omega t}$.

The instantaneous voltage and current equations may be obtained by multiplying Eqs. (10) and (11) by $e^{j\omega t}$ as indicated in Eqs. (1) and (2), thus

$$\mathbf{v} = A e^{j\omega t + \gamma x} + B e^{j\omega t - \gamma x} \tag{13}$$

$$i = \frac{A}{Z_0} e^{j\omega t + \gamma x} - \frac{B}{Z_0} e^{j\omega t - \gamma x} \tag{14}$$

These equations contain terms of the form $f_1(\omega t + \gamma x)$ and $f_2(\omega t - \gamma x)$, indicating the traveling-wave nature of the solution. In Fig. 1, the distance x is measured from the receiving end of the line. The terms containing $e^{j\omega t + \gamma x}$ represent waves traveling in the $-x$ direction and are therefore the *outgoing* waves of voltage and current. The terms containing $e^{j\omega t - \gamma x}$ represent waves traveling in the $+x$ direction and constitute the reflected waves of voltage and current. The ratio of voltage to current for either the outgoing or reflected wave is equal to the characteristic impedance of the line.

The constants A and B in the transmission-line equations may be evaluated in terms of known boundary conditions. Let us evaluate these in terms of the conditions at the receiving end of the line. At the receiving end we have $x = 0$, $V = V_R$, $I = I_R$ and $Z_R = V_R/I_R$. Equations (10) and (11) then become $V_R = A + B$ and $I_R = (1/Z_0)(A - B)$, from which we obtain

$$A = \frac{V_R}{2}\left(1 + \frac{Z_0}{Z_R}\right) \qquad B = \frac{V_R}{2}\left(1 - \frac{Z_0}{Z_R}\right) \tag{15}$$

Inserting these into Eqs. (10) and (11), and using $V_R/I_R = Z_R$, we obtain the transmission-line equations

$$V = \frac{V_R}{2}\left(1 + \frac{Z_0}{Z_R}\right)e^{\gamma x} + \frac{V_R}{2}\left(1 - \frac{Z_0}{Z_R}\right)e^{-\gamma x} \tag{16}$$

$$I = \frac{I_R}{2}\left(1 + \frac{Z_R}{Z_0}\right)e^{\gamma x} + \frac{I_R}{2}\left(1 - \frac{Z_R}{Z_0}\right)e^{-\gamma x} \tag{17}$$

The terms in Eqs. (16) and (17) may be regrouped to express these equations in hyperbolic function form. The hyperbolic functions are

$$\cosh \gamma x = \frac{e^{\gamma x} + e^{-\gamma x}}{2} \tag{18}$$

$$\sinh \gamma x = \frac{e^{\gamma x} - e^{-\gamma x}}{2} \tag{19}$$

$$\tanh \gamma x = \frac{e^{\gamma x} - e^{-\gamma x}}{e^{\gamma x} + e^{-\gamma x}} \tag{20}$$

In hyperbolic function form, Eqs. (16) and (17) become

$$V = V_R \left(\cosh \gamma x + \frac{Z_0}{Z_R} \sinh \gamma x \right) \tag{21}$$

$$I = I_R \left(\cosh \gamma x + \frac{Z_R}{Z_0} \sinh \gamma x \right) \tag{22}$$

$$Z = \frac{V}{I} = Z_0 \left(\frac{Z_R + Z_0 \tanh \gamma x}{Z_0 + Z_R \tanh \gamma x} \right) \tag{23}$$

The impedance Z is the ratio of voltage to current at any point on the line distant x from the receiving end. This is also the impedance which would be obtained if the line were cut at the point x and the impedance were measured looking toward the load.

8.03. Line Terminated in Its Characteristic Impedance.—If a transmission line is terminated in an impedance equal to its characteristic impedance, *i.e.*, if $Z_R = Z_0$, then the reflected-wave terms in Eqs. (8.02-16 and 17) vanish, leaving only the outgoing waves,

$$V = V_R e^{\gamma x} \tag{1}$$

$$I = I_R e^{\gamma x} \tag{2}$$

$$Z = \frac{V}{I} = Z_0 \tag{3}$$

Therefore, if the line is terminated in an impedance equal to its characteristic impedance, the impedance at any point on the line is equal to the characteristic impedance of the line. All of the energy in the outgoing wave is then absorbed in the terminating impedance and there is no reflection.

It is interesting to express Eqs. (1) and (2) in terms of the conditions at the sending end of the line. At the sending end we have $x = l$, $V = V_S$, and $I = I_S$. Equations (1) and (2) then become

$$V_S = V_R e^{\gamma l} \tag{4}$$

$$I_S = I_R e^{\gamma l} \tag{5}$$

Solving these for V_R and I_R and inserting these into Eqs. (1) and (2), with the additional substitution $s = l - x$, where s is the distance from the sending end to the point where V and I are taken, we have

$$V = V_S e^{-\gamma s} = V_S e^{-\alpha s} e^{-j\beta s} \tag{6}$$

$$I = I_S e^{-\gamma s} = I_S e^{-\alpha s} e^{-j\beta s} \tag{7}$$

The factor $e^{-\alpha s}$ in these equations signifies a decrease in magnitude or an attenuation of the outgoing wave as it travels toward the receiving end of the line. The attenuation constant α is given in nepers per unit length of line. The factor $e^{-j\beta s}$ denotes a phase shift of βs radians in the distance s, or β radians per unit length of line. Figure 3 shows the variation of the magnitude of the voltage or current as a function of distance for a line terminated in an impedance equal to its characteristic impedance.
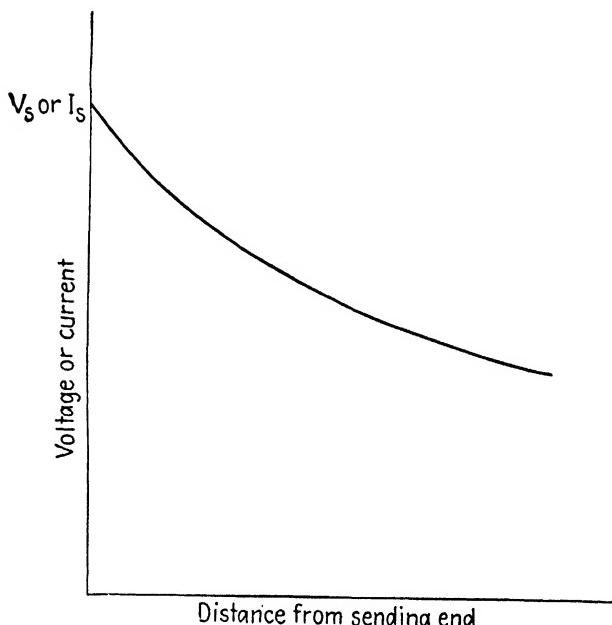
FIG. 3.—Magnitude of voltage and current as a function of distance along a line terminated in its characteristic impedance.

The voltages and currents in the above equations are the amplitudes or peak values of the sinusoidally varying functions. The scalar amplitudes are $|V| = V_S e^{-\alpha s}$ and $|I| = I_S e^{-\alpha s}$. For low-loss lines the voltage and current are in phase; the power flow at any point on a line terminated in its characteristic impedance is:

$$P = \tfrac{1}{2}|V||I|$$

$$= \frac{|V_S|^2}{2Z_0} e^{-2\alpha s} \tag{8}$$

The power loss per unit length of line is the space rate of decrease of the

transmitted power, or $P_L = -(dP/dx)$. Inserting Eq. (8), we obtain

$$P_L = 2\alpha \left(\frac{|V_S|^2}{2Z_0} e^{-2\alpha s} \right) = 2\alpha P$$

$$\alpha = \frac{P_L}{2P} \tag{9}$$

Consequently, the attenuation constant is the ratio of the power loss per unit length of line to twice the transmitted power.

8.04. Propagation Constant and Characteristic Impedance.—The propagation constant γ contains an attenuation constant α and a phase constant β. The phase constant represents the number of radians of phase shift per unit length of line. The *wavelength* λ is the distance required for a phase shift of 2π radians, or

$$\lambda = \frac{2\pi}{\beta} \tag{1}$$

The *phase velocity* v is the product of the frequency times wavelength, or

$$v = f\lambda = \frac{\omega}{\beta} \tag{2}$$

The propagation constant γ and characteristic impedance Z_0 are dependent upon the series impedance z and shunt admittance y per unit length of line as expressed by Eqs. (8.02-9 and 12).

For most transmission lines operating at frequencies above 100 kilocycles we find that $\omega L \gg R$ and $\omega C \gg G$, *i.e.*, the reactance and susceptance, are large in comparison with the resistance and conductance. We shall refer to such lines as low-loss lines. Simplified expressions may be derived for γ and Z_0 for this case. Expanding Eqs. (8.02-9 and 12) by the binomial series and retaining the first few terms of the series, we obtain

$$\gamma = \sqrt{zy} = (R + j\omega L)^{\frac{1}{2}}(G + j\omega C)^{\frac{1}{2}}$$

$$\approx \left(\frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}} \right) + j\omega\sqrt{LC} \tag{3}$$

$$Z_0 = \sqrt{\frac{z}{y}} = (R + j\omega L)^{\frac{1}{2}}(G + j\omega C)^{-\frac{1}{2}}$$

$$\approx \sqrt{\frac{L}{C}} \left[1 - j\left(\frac{R}{2\omega L} - \frac{G}{2\omega C} \right) \right] \tag{4}$$

Since we have assumed that $\omega L \gg R$ and $\omega C \gg G$, the characteristic impedance as given by Eq. (4), is substantially a pure resistance of value

$$Z_0 = \sqrt{\frac{L}{C}} \tag{5}$$

The real part of Eq. (3) is the attenuation constant, whereas the imaginary part is the phase constant. Inserting Eq. (5) into (3), we obtain

$$\alpha = \frac{R}{2Z_0} + \frac{GZ_0}{2} \tag{6}$$

$$\beta = \omega\sqrt{LC} \tag{7}$$

The wavelength and phase velocity for the low-loss line are

$$\lambda = \frac{2\pi}{\beta} = \frac{2\pi}{\omega\sqrt{LC}} \tag{8}$$

$$v = f\lambda = \frac{1}{\sqrt{LC}} \tag{9}$$

For a lossless line, *i.e.*, $R = G = 0$, the attenuation constant is zero and the outgoing and reflected waves experience no attenuation as they travel along the line. It can be shown that the phase velocity for a lossless line is equal to the velocity of light. The effect of losses in the line is to decrease both the wavelength λ and phase velocity v and to introduce attenuation in the line.

8.05. Transmission-line Parameters.—The R, L, C, and G parameters of several different types of transmission lines are given in Table 1. The resistance given in this table is the skin-effect resistance as computed by the methods of Chap. 15. The conductance is that resulting from dielectric losses in the insulating medium. The attenuation constants α_c and α_d are those resulting from losses in the conductor and dielectric, respectively. The total attenuation constant is the sum of the two terms, or $\alpha = \alpha_c + \alpha_d$. The attenuation constants and characteristic impedance equations are for low-loss lines.

The skin-effect resistance varies inversely as the radius of the conductor. In general, therefore, the attenuation constant decreases as the radius increases. A coaxial line has minimum attenuation for the ratio $b/a = 3.6$, corresponding to a characteristic impedance of approximately 77 ohms.

TABLE 1.—TRANSMISSION-LINE CONSTANTS
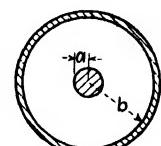
Inductance, L henrys per meter of line	$\dfrac{\mu_0}{\pi}\cosh^{-1}\dfrac{d}{2r}$ or $\dfrac{\mu_0}{\pi}\ln\dfrac{d}{r}$ if $d \gg r$	$\dfrac{\mu_0}{2\pi}\ln\dfrac{b}{a}$
Capacitance, C farads per meter of line	$\dfrac{\pi\epsilon}{\cosh^{-1}d/2r}$ or $\dfrac{\pi\epsilon}{\ln d/r}$ if $d \gg r$	$\dfrac{2\pi\epsilon}{\ln b/a}$
Resistance, R ohms per meter of line	$\dfrac{1}{r}\sqrt{\dfrac{\mu_0 f}{\pi\sigma}}$	$\left(\dfrac{1}{a}+\dfrac{1}{b}\right)\sqrt{\dfrac{\mu_0 f}{4\pi\sigma}}$
Conductance, G mhos per meter of line	ωDC	ωDC
Characteristic impedance in ohms $Z_0 = \sqrt{\dfrac{L}{C}}$	$120\cosh^{-1}\dfrac{d}{2r}$ or $276\log_{10}\dfrac{d}{r}$ if $d \gg r$	$\dfrac{138}{\sqrt{\epsilon_r}}\log_{10}\dfrac{b}{a}$
Attenuation constant due to conductor losses $\alpha_c = \dfrac{R}{2Z_0}$	$\dfrac{\sqrt{\mu_0 f/\pi\sigma}}{552\,r\log_{10}d/r}$	$\dfrac{[(1/a)+(1/b)]\sqrt{\mu_0\epsilon_r f/4\pi\sigma}}{276\log_{10}b/a}$
Attenuation constant due to dielectric losses $\alpha_d = \dfrac{GZ_0}{2}$	$\dfrac{GZ_0}{2}$	$\dfrac{GZ_0}{2}$

$\mu_0 = 4\pi \times 10^{-7}$ henry/meter

$\epsilon = \epsilon_0\epsilon_r$

$\epsilon_0 = 8.85 \times 10^{-12}$ farad/meter

ϵ_r = relative permittivity (dielectric constant)

D = dissipation factor of dielectric (for low-loss dielectrics, $D \approx$ P.F., where P.F is the power factor of the dielectric. See Appendix II for power factors of typical dielectrics.)

σ = conductivity of conductor

  = 5.80×10^7 mhos/meter for copper

  = 6.14×10^7 mhos/meter for silver

TABLE 2.

Transmission-line configuration	Characteristic impedance $Z_0 = \sqrt{\dfrac{L}{C}}$
	$Z_0 = 138 \log_{10} \dfrac{2h}{r}$
	$Z_0 = 276 \log_{10} \dfrac{d}{r}$
	$Z_0 = \dfrac{120}{\sqrt{\epsilon_r}} \left\{ \ln \left[2a \left(\dfrac{1 - b^2}{1 + b^2} \right) \right.\right.$ $\left.\left. - \dfrac{1 + 4a^2}{16a^4} (1 - 4b^2) \right] \right\}$ where $a = \dfrac{s}{2r} \qquad b = \dfrac{s}{d}$
	$Z_0 = \dfrac{138 \log_{10} b/a}{\sqrt{1 + (w/s)(\epsilon_r - 1)}}$ $s \ll \lambda$
	$Z_0 = \dfrac{276 \log_{10} b/a}{\sqrt{1 + (w/s)(\epsilon_r - 1)}}$ $s \ll \lambda$
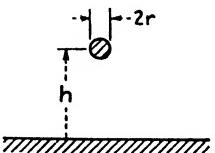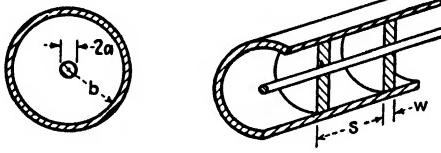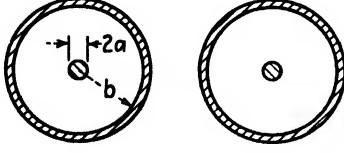
If a line is terminated by an impedance equal to its characteristic impedance, the decibel loss in the line is

$$\text{db} = 20 \log_{10} \frac{|V_S|}{|V_R|} = 20 \log_{10} \epsilon^{\alpha l} = 8.686 \alpha l \tag{1}$$

Table 2 gives the characteristic impedance of several of the more common types of transmission lines.

8.06. Lossless Line Equations.—In most microwave transmission lines we have $\omega L \gg R$ and $\omega C \gg G$ and, consequently, the transmission lines have characteristics approximating those of a lossless line. Let us therefore consider the transmission-line equations for the theoretical case of a lossless line.

A lossless line would have zero attenuation constant and hence $\gamma = j\beta$. Hyperbolic functions of imaginary angles may be written as trigonometric functions of real angles, that is

$$\sinh j\beta x = j \sin \beta x \qquad \tanh j\beta x = j \tan \beta x$$

$$\cosh j\beta x = \cos \beta x \qquad \coth j\beta x = -j \cot \beta x \tag{1}$$

For the lossless line, Eqs. (8.02-21, 22, and 23) become

$$V = V_R \left(\cos \beta x + j \frac{Z_0}{Z_R} \sin \beta x \right) \tag{2}$$

$$I = I_R \left(\cos \beta x + j \frac{Z_R}{Z_0} \sin \beta x \right) \tag{3}$$

$$Z = Z_0 \left(\frac{Z_R + jZ_0 \tan \beta x}{Z_0 + jZ_R \tan \beta x} \right) \tag{4}$$

Consider the case of a lossless line which is short-circuited at the distant end. We then have $Z_R = 0$ and $V_R = 0$. Remembering that $I_R = V_R/Z_R$, Eqs. (2), (3), and (4) become

$$V = jI_R Z_0 \sin \beta x \tag{5}$$

$$I = I_R \cos \beta x \tag{6}$$

$$Z = jZ_0 \tan \beta x \tag{7}$$

Equations (5) and (6) represent *standing waves* of voltage and current on the line as shown in Fig. 4. The standing wave is produced by a combination of an outgoing wave and a reflected wave, traveling in opposite directions on the line. Figure 4 may be visualized as representing the amplitudes of voltages and currents which have a sinusoidal time variation. The voltage is zero at the receiving end and has its maximum value at

points corresponding to $\beta x = n\pi/2$ or $x = n\lambda/4$, where n is any odd integer. The voltage and current are in space quadrature, as evidenced by Fig. 4, and also in time quadrature as indicated by the j term in Eq._(5).
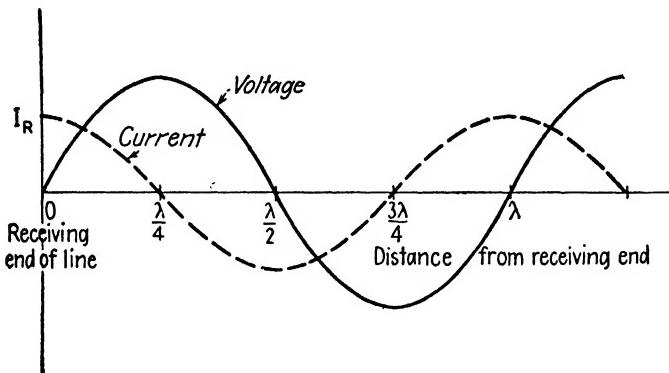
Fig. 4.—Standing waves of voltage and current on a short-circuited lossless line.

The impedance of the short-circuited lossless line, as given by Eq. (7), is plotted in Fig. 5. The input impedance is a pure reactance, alternating between capacitive and inductive reactance as βx increases. Antiresonance occurs when the line is an odd integral number of quarter wavelengths long and resonance occurs when it is an even integral number of quarter wavelengths long. The antiresonant input impedance of a lossless line
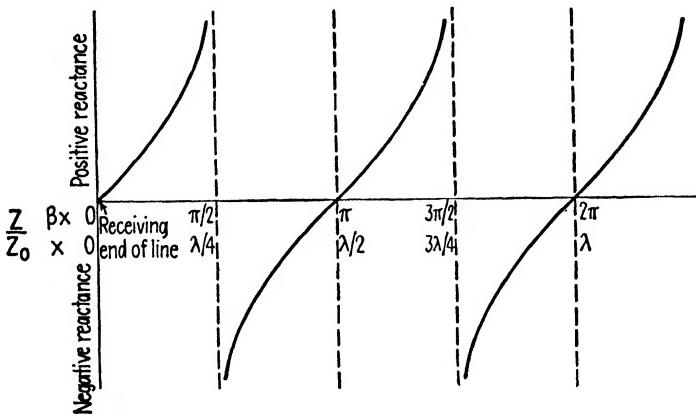
Fig. 5.—Impedance ratio Z/Z_0 for a short-circuited lossless line as a function of x and βx.

would be a pure resistance of theoretically infinite value, whereas the resonant impedance would be zero. In the practical case of a low-loss line, the impedance is a very large pure resistance for antiresonance and a very small pure resistance for resonance.

Now consider the lossless line which is open-circuited at the distant end. For this case we have $Z_R = \infty$, $I_R = 0$, and $V_R = I_R Z_R$. Equations (2), (3), and (4) then become

$$V = V_R \cos \beta x \tag{8}$$

$$I = j\frac{V_R}{Z_0} \sin \beta x \tag{9}$$

$$Z = -jZ_0 \cot \beta x \tag{10}$$

The voltage and current standing waves are similar to those shown in Fig. 4, but with voltage and current interchanged. The impedance is a
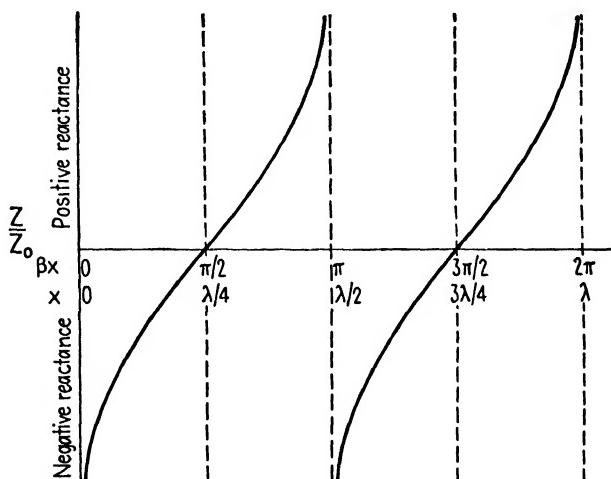
FIG. 6.—Impedance ratio Z/Z_0 for an open-circuited lossless line as a function of x and βx.

pure reactance as shown in Fig. 6. Resonance occurs when the line is an odd integral number of quarter wavelengths long and antiresonance when it is an even integral number of quarter wavelengths long.

We can now conclude that the short-circuited and open-circuited lines are either resonant or antiresonant when the length l is $l = n\lambda/4$ or when $\beta l = n\pi/2$, where n is given by

	Short-circuited line	Open-circuited line
Resonance...............	n = even integer	n = odd integer
Antiresonance	n = odd integer	n = even integer

8.07. Short-circuited Line with Losses.—If the line losses are not zero and the line is short-circuited at the distant end, we again have $Z_R = 0$, $V_R = 0$, and $I_R = V_R/Z_R$. Equations (8.02-21, 22, and 23) then reduce to

$$V = I_R Z_0 \sinh \gamma x \qquad (1)$$

$$I = I_R \cosh \gamma x \qquad (2)$$

$$Z = Z_0 \tanh \gamma x \qquad (3)$$
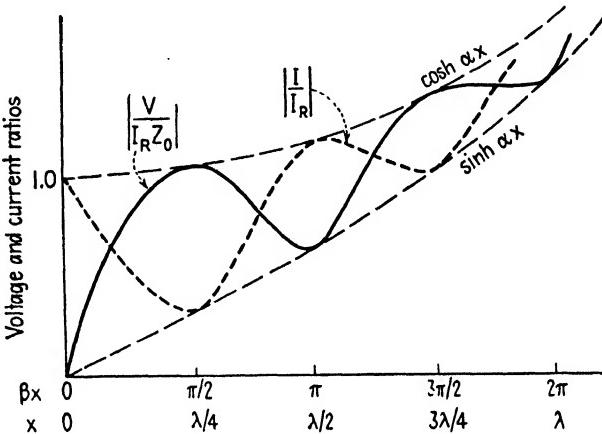
FIG. 7.—Voltage ratio $\left| \dfrac{V}{I_R Z_0} \right|$ and current ratio $\left| \dfrac{I}{I_R} \right|$ as a function of x and βx for a short-circuited line having losses.

Inserting $\gamma = \alpha + j\beta$ into these equations and applying the identities for the hyperbolic function of the sum of two angles,[1] we obtain

$$\frac{V}{I_R Z_0} = \sinh \alpha x \cos \beta x + j \cosh \alpha x \sin \beta x \qquad (4)$$

$$\frac{I}{I_R} = \cosh \alpha x \cos \beta x + j \sinh \alpha x \sin \beta x \qquad (5)$$

$$\frac{Z}{Z_0} = \frac{\tanh \alpha x + j \tan \beta x}{1 + j \tanh \alpha x \tan \beta x} \qquad (6)$$

The scalar values of $\left| V/I_R Z_0 \right|$ and $\left| I/I_R \right|$ are plotted against x and βx in Fig. 7. Equation (4) shows that when $\beta x = n\pi/2$, the ratio $\left| V/I_R Z_0 \right|$ has the value $\cosh \alpha x$ or $\sinh \alpha x$, depending upon whether n is an odd or even integer. Likewise, Eq. (5) shows that when $\beta x = n\pi/2$, the current ratio I/I_R has the values $\cosh \alpha x$ and $\sinh \alpha x$ for even and odd integer

[1] See, for example, B. O. PIERCE, "A Short Table of Integrals," Ginn and Company Boston, 1929.

values of n, respectively. The curves $\cosh \alpha x$ and $\sinh \alpha x$ therefore represent the envelope of the curves $\mid V/I_R Z_0 \mid$ and $\mid I/I_R \mid$, as shown in **Fig. 7**. Since I_R and Z_0 are independent of x the ratio $\mid V/I_R Z_0 \mid$ represents the voltage distribution along the line, and the ratio $\mid I/I_R \mid$ represents the current distribution.

Referring to Eq. (6), we find that the scalar impedance ratio $\mid Z/Z_0 \mid$ varies between the limits $\tanh \alpha x$ and $\coth \alpha x$ as shown in **Fig. 8**. At the
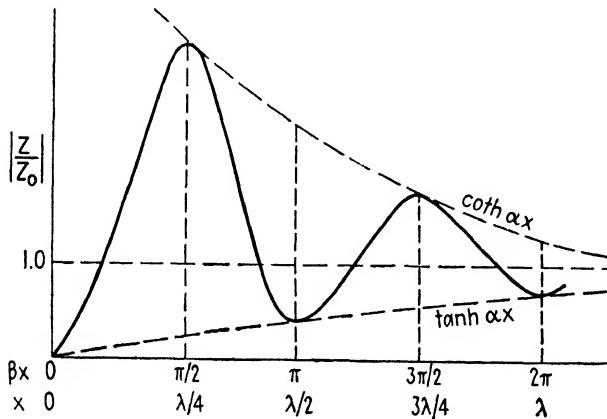
FIG. 8.—Magnitude of the impedance ratio Z/Z_0 for a short-circuited line having losses.

points where the impedance has its maximum and minimum values, the impedances are approximately $Z = Z_0 \coth \alpha x$ and $Z = Z_0 \tanh \alpha x$, respectively.

Figures 7 and 8 represent scalar values, hence all values are plotted as positive quantities. For the lossless line, the curves in Figs. 7 and 8 would degenerate to curves similar to those in Figs. 4 and 5 but with all values plotted as positive quantities.

The variable βx in the above figures may be written $\beta x = \omega x/v$. Consequently, we may consider the above curves as being plotted either against frequency, with line length held constant, or against length of line, with frequency held constant.

8.08. Receiving End Open-circuited.—If the receiving end is open-circuited, we have $Z_R = \infty$, $I_R = 0$; hence Eqs. (8.02-21, 22, 23) yield

$$V = V_R \cosh \gamma x \tag{1}$$

$$I = \frac{V_R}{Z_0} \sinh \gamma x \tag{2}$$

$$Z = Z_0 \coth \gamma x \tag{3}$$

Comparison of these equations with Eqs. (8.07-1, 2, and 3) shows that the ratio $\mid V/V_R \mid$ for the open-circuited line has a variation with γx similar

to the current ratio of the short-circuited line, whereas the ratio $|IZ_0/V_R|$ of the open-circuited line is similar to the voltage ratio of the short-circuited line. The impedance ratio $|Z/Z_0|$ for the open-circuited line is
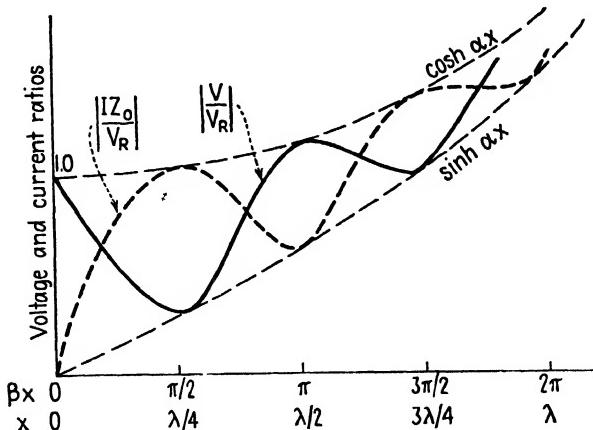
FIG. 9.—Voltage and current ratios for an open-circuited line having losses.
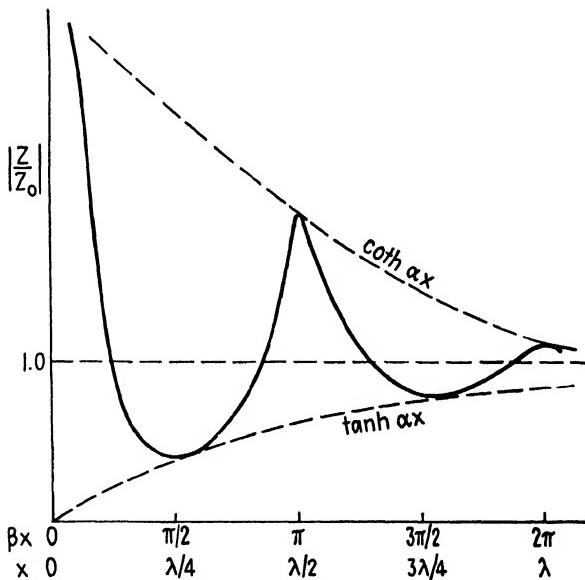
FIG. 10.—Magnitude of the impedance ratio for an open-circuited line having losses.

the reciprocal of that of the short-circuited line. These ratios are shown in Figs. 9 and 10.

8.09. Sending-end Equations.—Thus far we have dealt largely with the transmission-line equations expressed in terms of the voltage, current, and

impedance at the receiving end of the line. If the sending-end voltage V_s and current I_s are known, we may readily obtain V_R and I_R by writing Eqs. (8.02-16 and 17) or (8.02-21 and 22) for the full length of the line (letting $x = l$) and substituting the known values of V_s and I_s for V and I. These equations may then be solved for V_R and I_R.

If only the generated voltage V_g of Fig. 1 is known, it is then necessary to first compute the input impedance of the line by Eq. (8.02-23). The impedance as seen by the generator voltage V_g is the sum of the input impedance to the line and the generator impedance Z_g. The sending-end voltage and current may therefore be obtained from

$$I_s = \frac{V_g}{Z + Z_g} \tag{1}$$

$$V_s = V_g - I_s Z_g \tag{2}$$

where Z is the input impedance of the line.

PROBLEMS

1. Show that the binomial expansion of the terms given in Eqs. (8.04-3 and 4) yields the approximations indicated in these equations.
2. A coaxial line has dimensions $a = 0.75$ cm and $b = 3$ cm. The dielectric is polystyrene with a dielectric constant of 2.5 and power factor 0.0004. Compute the following values at a frequency of 500 megacycles:
   (a) Inductance and capacitance per meter of line.
   (b) Conductance and skin-effect resistance per meter of line.
   (c) Attenuation constant and phase constant.
   (d) Wavelength and phase velocity.
   (e) Input impedance of a quarter-wavelength section of line if (1) short-circuited and (2) open-circuited.
3. A lossless line is one-eighth of a wavelength long and is terminated by a pure resistance which is approximately equal to the characteristic impedance of the line. Show that if the value of the terminating resistance is varied by a small amount either side of the value $R_L = Z_0$, the input impedance of the line will contain a reactive component, the magnitude of which varies directly with R, whereas the resistive component remains substantially constant.
4. A high-frequency voltmeter is constructed of a quarter-wavelength lossless line which is terminated in the heater junction of a thermocouple having a resistance of R ohms. The thermocouple leads are connected to a microammeter.
   (a) Derive an expression for the voltage V_s at the sending end in terms of the current I_R through the thermocouple heater.
   (b) Derive an equation for the input impedance to the line. How does this vary with the value of the thermocouple resistance?
   (c) Compute the input voltage and input impedance if $R = 5$ ohms, $Z_0 = 75$ ohms and $I_R = 15$ ma.

CHAPTER 9

GRAPHICAL SOLUTION OF TRANSMISSION-LINE PROBLEMS

A number of ingenious circle diagrams have been devised to facilitate the graphical solution of transmission-line problems. Basically, all of these spring from the same fundamental relationships which are expressed in the transmission-line equations. In this chapter, we shall consider two types of impedance diagrams, these being referred to as the *rectangular impedance diagram* and the *polar impedance diagram*. First, however, let us derive the transmission-line equations in reflection-coefficient form, since these will be useful in the construction of the impedance diagrams.

9.01. Reflection-coefficient Equations.—The reflection coefficient r_R is defined as the ratio of the reflected voltage to the outgoing voltage at the receiving end of the line. In Eq. (8.02-15) the terms A and B represent the outgoing and reflected voltages at the load, respectively. The reflection coefficient is therefore

$$r_R = \frac{Z_R - Z_0}{Z_R + Z_0} \tag{1}$$

Let us now express the transmission-line equations in reflection-coefficient form. In Eqs. (8.02-16 and 17), let $V'_R = V_R/2[1 + (Z_0/Z_R)]$, where V'_R is the outgoing-voltage wave at the load. Now factor out the term $V'_R e^{\gamma x}$ and, with the substitution of r_R from Eq. (1), we obtain

$$V = V'_R e^{\gamma x}(1 + r_R e^{-2\gamma x}) \tag{2}$$

$$I = \frac{V'_R}{Z_0} e^{\gamma x}(1 - r_R e^{-2\gamma x}) \tag{3}$$

$$Z = \frac{V}{I} = Z_0 \left(\frac{1 + r_R e^{-2\gamma x}}{1 - r_R e^{-2\gamma x}} \right) \tag{4}$$

Equations (2), (3), and (4) are the transmission-line equations expressed in terms of the reflection coefficient. They may be used to evaluate the voltage, current, and impedance at any point on the line. The terms $V'_R e^{\gamma x}$ and $(V'_R/Z_0)e^{\gamma x}$ in these equations represent the outgoing waves of voltage and current, respectively, whereas the terms $r_R V'_R e^{-\gamma x}$ and $-r_R(V'_R/Z_0)e^{-\gamma x}$ represent the reflected waves.

If the line is terminated in an impedance equal to its characteristic impedance, Eq. (1) shows that the reflection coefficient is zero and conse-

quently the reflected-wave terms in Eqs. (2) and (3) are zero, leaving only the outgoing waves.

If the line is short-circuited at the distant end, we have $Z_R = 0$ and $r_R = -1$; for an open-circuited line, we have $Z_R = \infty$ and $r_R = 1$. In the more general case, r_R is complex and may be readily evaluated using Eq. (1).

9.02. The Rectangular Impedance Diagram.[1]—In the construction of impedance diagrams, it is convenient to express the reflection coefficient as an exponential quantity. We therefore let

$$r_R = e^{-2(t_0 + j u_0)} \tag{1}$$

where e^{-2t_0} is the magnitude and $-2u_0$ is the angle of the reflection coefficient. Now let

$$t = t_0 + \alpha x \tag{2}$$

$$u = u_0 + \beta x \tag{3}$$

In Eq. (9.01-4) the term $r_R e^{-2\gamma x}$ then becomes $r_R e^{-2\gamma x} = e^{-2(t + ju)}$ and the impedance ratio may be written:

$$\frac{Z}{Z_0} = \frac{1 + e^{-2(t+ju)}}{1 - e^{-2(t+ju)}} \tag{4}$$

At the receiving end of the line, we have $Z = Z_R$, $\alpha x = 0$, and $\beta x = 0$; therefore Eq. (4) becomes

$$\frac{Z_R}{Z_0} = \frac{1 + e^{-2(t_0 + j u_0)}}{1 - e^{-2(t_0 + j u_0)}} \tag{5}$$

Equation (4) is the basic relationship used in the construction of impedance diagrams. The impedance diagram is essentially a plot of Eq. (4) which enables us to obtain values of Z/Z_0 if the values of t and u are known, or conversely, to obtain values of t and u if Z/Z_0 is known. When impedances are expressed as a ratio, such as Z/Z_0, they are known as *normalized impedances*.

For convenience, let us separate the real and imaginary parts of Z/Z_0 in Eq. (4), letting [2]

$$\frac{Z}{Z_0} = r + jx = \frac{1 + e^{-2(t+ju)}}{1 - e^{-2(t+ju)}} \tag{6}$$

[1] The theory of the impedance diagram presented here is similar to that given by W. JACKSON and L. G. HUXLEY in The Solution of Transmission-line Problems by the Use of the Circle Diagram of Impedance, *J.I.E.E.* (*London*), vol. 91, part 3, pp. 105–127; September, 1944.

[2] For low-loss lines it may be assumed that Z_0 is real. Writing $Z = R + jX$, we obtain $Z/Z_0 = (R/Z_0) + j(X/Z_0)$. Here $r = R/Z_0$ corresponds to the normalized resistance and $x = X/Z_0$ is the normalized reactance.

The rectangular impedance diagram is a plot of Eq. (6) in the form of constant-t and constant-u loci in the $r + jx$ plane as shown in Fig. 1. If t is held constant in Eq. (6) and u is allowed to vary, the locus of r and x for various values of u may be shown to be a circle with its center on the r axis, distant coth $2t$ from the origin, and with a radius cosech $2t$. On the
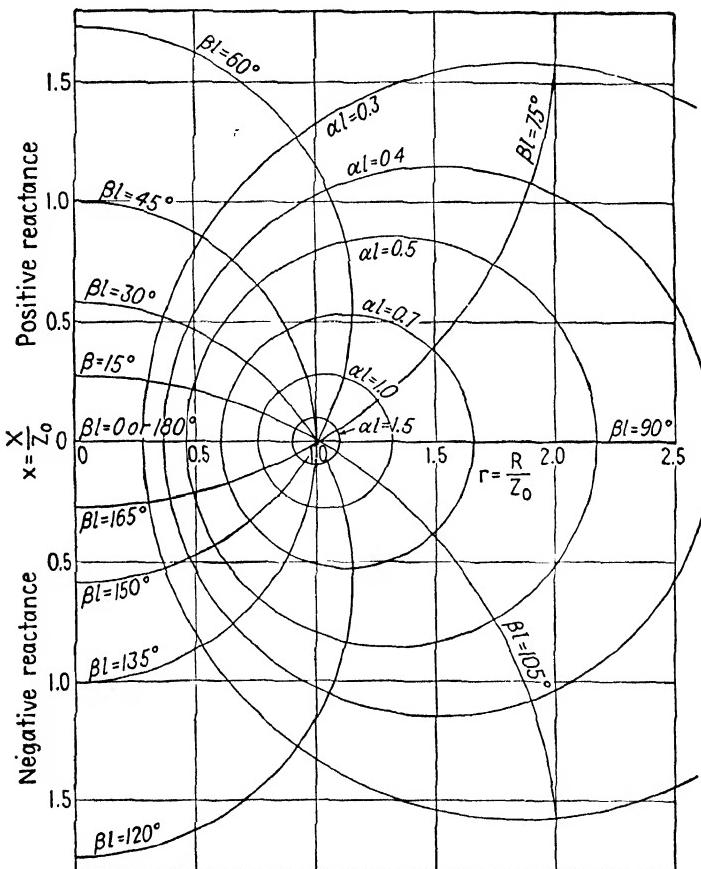
Fig. 1.—Rectangular impedance diagram.

other hand, if u is held constant and t is allowed to vary, the locus of r and x is a circle centered on the x axis, distant $-$ cot $2u$ from the origin, and with a radius cosec $2u$. Since t is related to αl by Eq. (2) and u is related to βl by Eq. (3), it is customary to designate the constant-t and constant-u circles as αl and βl circles, respectively, as shown in Fig. 1. These constitute two families of orthogonal bipolar circles.[1] Any point in

[1] Bipolar circles of the type shown in Fig. 1 are encountered in a number of engineering applications. For example, the αl and βl circles correspond to the electric and magnetic field lines of a two-conductor transmission line.

the impedance diagram represents a value of $r + jx$ and a value of $t + ju$ (read on the αl and βl circles), these two quantities being related by Eq. (6).

Let us now trace the steps which are necessary to evaluate the input impedance of a transmission line terminated in a known impedance Z_R.

It is assumed that Z_R, Z_0, α, β, and the length of line l are known. The procedure is as follows:

1. Compute the values of Z_R/Z_0, αl, and βl.

2. Enter the chart at the known value of Z_R/Z_0 and observe the corresponding values of t_0 and u_0 (on the αl and βl circles).

3. Compute the values of $t = t_0 + \alpha l$ and $u = u_0 + \beta l$. These are the values of t and u corresponding to the sending end of the line.

4. Reenter the chart at the new values t and u (on the αl and βl circles) and read the corresponding impedance Z_l/Z_0. This is the normalized input impedance.

As an example, assume that $Z_R/Z_0 = 2 + j0$, $\alpha l = 0.2$ nepers, and $\beta l = 0.6$ radians. Entering the impedance diagram of Fig. 1 at $Z_R/Z_0 = 2 + j0$, we obtain the values of $t_0 = 0.534$ and $u_0 = 90°$ or 1.57 radians. Step 3 above yields $t = 0.734$ and $u = 124°$ or 2.17 radians. Reentering the impedance diagram at these values of t and u, the normalized input impedance is found to be $Z/Z_0 = 1.08 - j0.5$.

9.03. Polar Impedance Diagram.—In the rectangular impedance diagram, the circle $\alpha l = 0$ has an infinitely large radius. Therefore an infinitely large diagram would be required to solve all possible problems. When dealing with low-loss lines which are open-circuited, short-circuited, or terminated in a pure reactance, the solution is sometimes found to be beyond the limits of any practical diagram. In the polar impedance diagram, introduced by P. H. Smith,[1, 2] the entire impedance diagram is contained within a circle of any desired radius.

The general plan of construction of the polar impedance diagram will be given here, followed by several illustrative problems. A more detailed description of the construction of the diagram is given in Sec. 9.07.

In the polar impedance diagram, the impedance $Z/Z_0 = r + jx$ and the quantity $t + ju$ in Eq. (9.02-6) are both related to a new variable $p + jq$. Let

$$p + jq = e^{-2(t+ju)} \tag{1}$$

and Eq. (9.02-6) then becomes

$$\frac{Z}{Z_0} = r + jx = \frac{1 + (p + jq)}{1 - (p + jq)} \tag{2}$$

---

[1] SMITH, P. H., Transmission-Line Calculator, *Electronics*, vol. 12, pp. 29–31; January, 1939.

[2] SMITH, P. H., An Improved Transmission-Line Calculator, *Electronics*, vol. 17, p. 130; January, 1944.
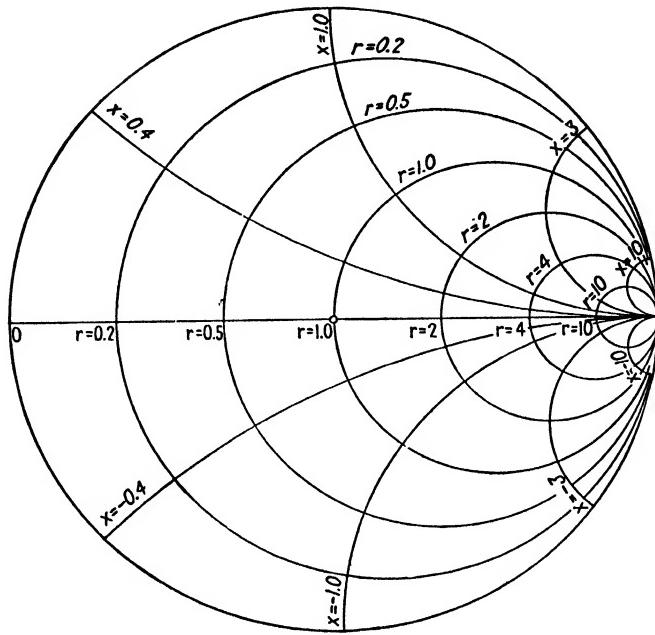
(a)
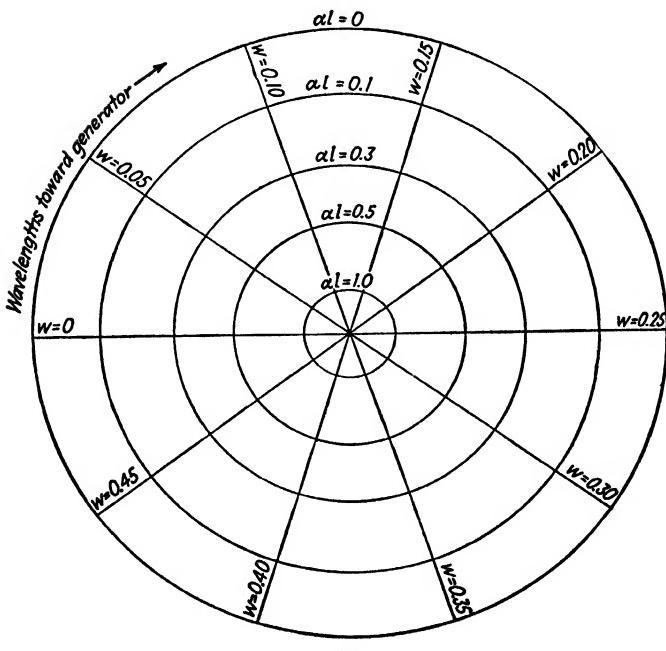
(b)

FIG. 2.—Polar impedance diagram. (a) Constant-*r* and constant-*x* loci; (b) constant-*t* (αl, and constant-*w* loci.

Let p and q represent the rectangular coordinate axes. Equation (2) may be used to obtain families of constant-r and constant-x loci in the $p + jq$ plane. Similarly, Eq. (1) may be used to plot families of constant-t and constant-u loci in the $p + jq$ plane. Any point in the polar impedance diagram then defines three quantities, (1) a value of $r + jx$, (2) a value of $t + ju$, and (3) a value of $p + jq$. These three quantities are interrelated by Eqs. (1) and (2). In the solution of transmission-line problems, we are interested in obtaining values of $r + jx$ corresponding to known values of $t + ju$, or vice versa. Once the diagram has been constructed, the p and q coordinate axes have no further use and, therefore, are omitted in the final impedance diagram.

The constant-r and constant-x loci form two families of orthogonal circles in the $p + jq$ plane as shown in Fig. 2a. The constant-r circles have centers on the p axis, distant $r/(1 + r)$ from the origin, and have a radius of $1/(r + 1)$. The entire impedance diagram is contained in a circle of unit radius, with center at the origin. The constant-x circles have centers at $p = 1$, $q = 1/x$ and have radii $1/x$. The upper half of the diagram of Fig. 2a represents positive reactance, whereas the lower half represents negative reactance. The constant-r and constant-x circles all pass through the point $(1, 0)$.

The constant-t loci consist of a family of circles having centers at the origin and radii e^{-2t}. These are designated [1] al in Fig. 2b. The constant-u loci are radial lines passing through the origin. However, it is more convenient to replace u by a new variable $w = u/2\pi$. Substituting u from Eq. (9.02-3), with $\beta = 2\pi/\lambda$ and $w_0 = u_0/2\pi$, we obtain

$$w = w_0 + \frac{x}{\lambda} \tag{3}$$

Therefore, w is a measure of the length of line in wavelengths. The constant-w lines are also radial lines passing through the origin, with a slope $- \tan 4\pi w$ as shown in Fig. 2b.

The polar impedance diagram of Fig. 3 is obtained by superimposing Figs. 2a and 2b. To simplify the final diagram, the constant-w lines have been omitted, although the values are given on the scales marked "wavelengths toward the generator" and "wavelengths toward the load" along the outer rim of the diagram.

[1] The impedance diagram of Fig. 3 contains constant-l circles rather than the

It is interesting to observe that both the rectangular and polar diagrams may be used to solve problems in terms of admittances as well as in terms of impedances. The normalized admittance at any point on the line is the reciprocal of the normalized impedance, thus Eq. (9.02-4) may be written

$$\frac{Y}{Y_0} = \frac{1 - e^{-2(t+ju)}}{1 + e^{-2(t+ju)}} \tag{4}$$

where $Y_0 = 1/Z_0$ is the characteristic admittance of the line. Equation (4) gives the same families of curves in the rectangular and polar diagrams as Eq. (9.02-4). The constant-r and constant-x circles become constant-conductance and constant-susceptance circles, respectively, when dealing with admittances. Otherwise, the use of the diagram is exactly the same for either impedances or admittances.

9.04. Use of the Polar Impedance Diagram.—The polar impedance diagram is used in much the same manner as the rectangular impedance diagram. To determine the input impedance of a line terminated in a known impedance, the procedure is as follows:

1. Compute Z_R/Z_0, αl, and l/λ.
2. Enter the impedance diagram at the known value of Z_R/Z_0 and read the corresponding values of t_0 on the αl circles and w_0 on the "wavelengths toward the generator" scale.
3. Compute the values of $t = t_0 + \alpha l$ and $w = w_0 + (l/\lambda)$.
4. Reenter the diagram at the new values of t and w and read the corresponding normalized sending-end impedance.

Referring to Fig. 4, assume that point P corresponds to the terminal impedance Z_R/Z_0 and that the line is lossless. As we move toward the generator on the transmission line, the impedance point moves in the clockwise direction on the constant-αl circle along the path PQ_1. However, if the line has losses, then the impedance point spirals inward as indicated by the path PQ_2. As the length of the line increases, the impedance point continues to spiral inward, eventually winding up on the point $Z/Z_0 = 1$.

If we move toward the generator on the transmission line, the impedance point moves in the clockwise direction in the impedance diagram and the values of w are read on the "wavelengths toward the generator" scale. If we move toward the load, the impedance point in the diagram moves in the counterclockwise direction and the values of w are read on the "wavelengths toward the load" scale. One complete revolution on the impedance diagram corresponds to a half wavelength of line.

9.05. Standing-wave Ratio.—The standing-wave ratio for a lossless line is defined as $\rho = \dfrac{|V_{max}|}{|V_{min}|}$, where $|V_{max}|$ and $|V_{min}|$ are the magnitudes of the maximum and minimum voltages. The standing-wave ratio is of

considerable importance, since it is a quantity which can readily be computed from laboratory measurements.

To obtain expressions for the maximum and minimum voltages, we return to Eq. (9.01-2) and substitute Eq. (9.02-1) for r_R. For a lossless line, $\alpha = 0$ and we have

$$V = V_R' e^{j\beta x}[1 + e^{-2t_0}e^{-2j(u_0+\beta x)}] \tag{1}$$
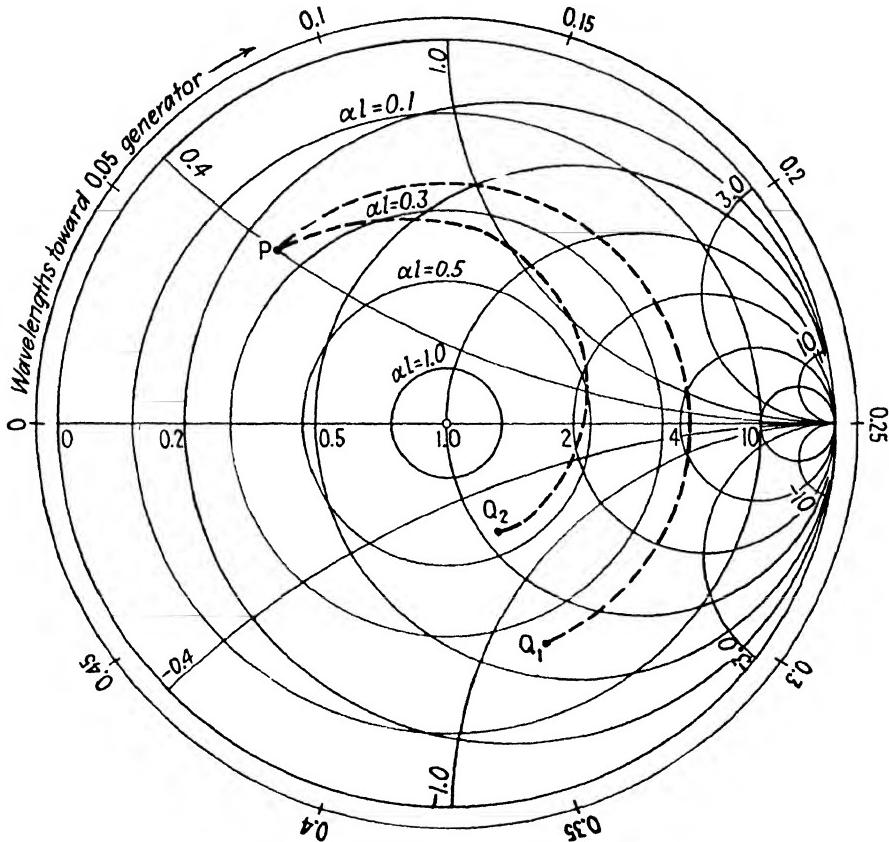
Fig. 4.— Use of the polar impedance diagram.

Maximum voltage occurs at that point on the line which makes the second term inside the bracket in phase with the first term. This occurs at that point on the line where $e^{-2j(u_0+\beta x_{\max})} = 1$, yielding

$$|V_{\max}| = V_R'(1 + e^{-2t_0}) \tag{2}$$

Minimum voltage occurs when $e^{-2j(u_0+\beta x_{\min})} = -1$, yielding

$$|V_{\min}| = V_R'(1 - e^{-2t_0}) \tag{3}$$

The standing-wave ratio is therefore

$$\rho = \left| \frac{V_{\max}}{V_{\min}} \right| = \frac{1 + e^{-2t_0}}{1 - e^{-2t_0}} = \frac{1 + |r_R|}{1 - |r_R|} \tag{4}$$

where $|r_R| = e^{-2t_0}$ is the scalar magnitude of the reflection coefficient.

Figure 5 shows a graph of the standing-wave ratio as a function of t_0. If the normalized impedance Z_R/Z_0 is known, the value of t_0 may be obtained from Fig. 3, and Fig. 5 may then be used to determine the standing-wave ratio.
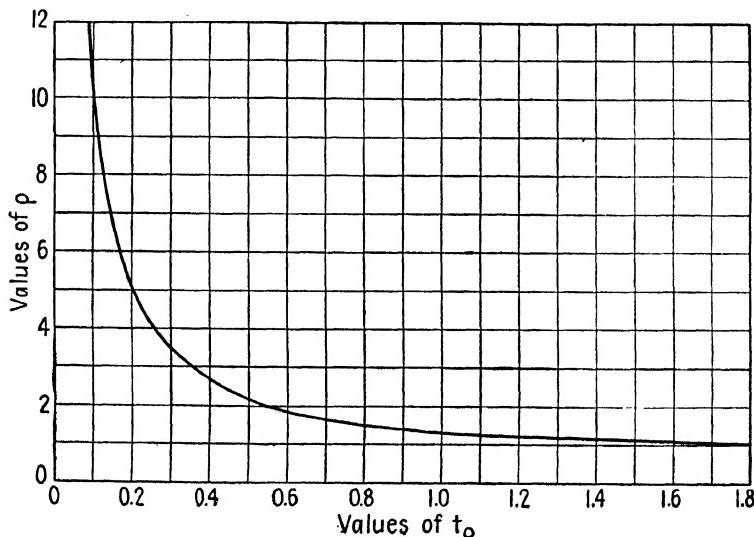
FIG. 5.—Standing-wave ratio ρ as a function of t_0.

The maximum and minimum voltages, respectively, occur at points on the line where

$$2(u_0 + \beta x_{\max}) = n\pi \qquad n \text{ is even} \tag{5}$$

$$2(u_0 + \beta x_{\min}) = n\pi \qquad n \text{ is odd} \tag{6}$$

These may be expressed as

$$\frac{x_{\max}}{\lambda} = \frac{n}{4} - \frac{u_0}{2\pi} = \frac{n}{4} - w_0 \qquad n \text{ is even} \tag{7}$$

$$\frac{x_{\min}}{\lambda} = \frac{n}{4} - \frac{u_0}{2\pi} = \frac{n}{4} - w_0 \qquad n \text{ is odd} \tag{8}$$

where $-2u_0$ is the phase angle of the reflection coefficient as given by Eq. (9.02-1).

If a line has attenuation, the successive values of $|V_{\max}|$ and $|V_{\min}|$

vary along the line and consequently there is no fixed value for the standing wave ratio.

9.06. Illustrative Examples.—The use of the polar impedance diagram, will now be illustrated by several examples.

Example 1.   Determine the input impedance of a line having the following characteristics:

$$Z_0 = 75 \text{ ohms} \qquad \frac{l}{\lambda} = 0.2$$

$$Z_R = 150 + j100 \text{ ohms} \qquad \alpha l = 0.15 \text{ neper}$$

We first compute $Z_R/Z_0 = 2 + j1.333$. Entering the polar impedance diagram at this value of normalized impedance, we obtain $t_0 = 0.35$ and $w_0 = 0.210$ (on the "wavelengths toward the generator" scale). Now compute $t = t_0 + \alpha l = 0.50$ and $w = w_0 + (l/\lambda) = 0.410$. These are the values of t and w at the sending end of the line. Reentering the impedance diagram at these values of t and w, we obtain the normalized input impedance $Z/Z_0 = 0.60 - j0.46$, or $Z = 45 - j34.5$ ohms. If the line were lossless the standing-wave ratio from Fig. 5, corresponding to $t_0 = 0.350$ would be $\rho = 2.97$.

Example 2.   A lossless line is terminated in a capacitance such that $Z_R/Z_0 = 0 - j0.5$. Determine the lengths of line required for (a) resonance ($Z/Z_0 = 0$) and (b) antiresonance ($Z/Z_0 = \infty$).

The outer circle of the impedance diagram corresponds to zero resistance. Entering the diagram at $Z_R/Z_0 = 0 - j0.5$, we obtain $t_0 = 0$ and $w_0 = 0.425$. The lengths of line required for the impedance point to move to (a) $Z/Z_0 = 0$ and (b) $Z/Z_0 = \infty$ are

   (a) $l = (0.5 - 0.425)\lambda = 0.075\lambda$.

   (b) $l = (0.75 - 0.425)\lambda = 0.325\lambda$.

The antiresonant line is a quarter wavelength longer than the resonant line.

Example 3.   A line which is 0.4λ long is short-circuited at one end and has a lumped impedance (normalized) of $Z_1/Z_0 = 0.5 + j0.2$ shunted across the input terminals. The value of αl is 0.2 nepers. Find the normalized input impedance.

In dealing with parallel circuits, it is advisable to use admittances. The impedance diagram may be used to convert impedances into admittances and vice versa. To determine the normalized admittance Y_1/Y_0, enter the impedance diagram at the point corresponding to the normalized impedance $Z_1/Z_0 = 0.5 + j0.2$. The normalized admittance is halfway around the diagram on the same constant-αl circle (or on a straight line through the center of the impedance diagram to the same constant-αl circle). This gives $Y_1/Y_0 = 1.72 - j0.72$.

Now consider the input admittance to the line only (with Y_1 disconnected). The admittance of the short circuit is $Y_R = \infty$ or $Y_R/Y_0 = \infty$. Entering the diagram at this value of admittance, we read $t_0 = 0$ and $w_0 = 0.25$. At the sending end of the line, we have $t = t_0 + \alpha l = 0.2$ and $w = w_0 + (l/\lambda) = 0.65$. The w scale returns to zero at $w = 0.5$; hence the point corresponding to $w = 0.65$ is 0.15 beyond the zero point (on the "wavelengths toward the generator" scale). The normalized input admittance of the line is, therefore, $Y_L/Y_0 = 0.54 + j1.23$. The normalized input admittance with Y_1 connected is then $Y/Y_0 = (Y_L/Y_0) + (Y_1/Y_0) = 2.26 + j0.51$. The normalized impedance is obtained by entering the diagram at $Y/Y_0 = 2.26 + j0.51$ and proceeding halfway around the diagram. Thus the input impedance is $Z/Z_0 = 0.43 - j0.09$.

9.07. Construction of the Polar Impedance Diagram.—The polar diagram consists of families of constant r, x, t, and w loci plotted in the $p + jq$ plane. In order to construct the diagram, it is first necessary to obtain

expressions relating each of the quantities r, x, t, and w, to p and q. These equations determine the respective loci.

Returning to Eq. (9.03-2), we add 1.0 to both sides of the equation, yielding

$$(r + 1) + jx = \frac{2}{1 - (p + jq)} \tag{1}$$

To separate the real and imaginary parts, multiply the numerator and denominator of the right-hand side by the conjugate of the denominator, yielding

$$(r + 1) + jx = \frac{2(1 - p)}{(1 - p)^2 + q^2} + \frac{j2q}{(1 - p)^2 + q^2} \tag{2}$$

Equating the real and imaginary terms on both sides of Eq. (2), we obtain

$$r + 1 = \frac{2(1 - p)}{(1 - p)^2 + q^2} \tag{3}$$

$$x = \frac{2q}{(1 - p)^2 + q^2} \tag{4}$$

Equation (3) may be written as

$$p^2 - \frac{2pr}{r + 1} + q^2 = -\frac{r - 1}{r + 1} \tag{5}$$

Adding $r^2/(r + 1)^2$ to both sides of the equation to complete the square, gives

$$\left(p - \frac{r}{r + 1}\right)^2 + q^2 = \frac{1}{(r + 1)^2} \tag{6}$$

This is the equation of the constant-r circles in Figs. 2a and 3. The centers of the circles are on the p axis at $r/(r + 1)$ and the radii are $1/(r + 1)$. The circle corresponding to $r = 0$ has unit radius and is centered at the origin in the $p + jq$ plane. The circle $r = \infty$ has zero radius and is centered at $p = 1$, $q = 0$.

To obtain the equation expressing the loci of the constant-x circles, Eq. (4) is rearranged to give

$$(p - 1)^2 + q^2 - \frac{2q}{x} = 0 \tag{7}$$

Adding $1/x^2$ to both sides of the equation completes the square and gives

$$(p - 1)^2 + \left(q - \frac{1}{x}\right)^2 = \frac{1}{x^2} \tag{8}$$

This is the equation of the constant-x circles, with centers at $p = 1$, $q = 1/x$, and radii $1/x$. The circle $x = \infty$ has zero radius.

To obtain the constant-t and constant-w circles, we return to Eq. (9.03-1) and write this in the form

$$p + jq = e^{-2t}(\cos 2u - j \sin 2u) \tag{9}$$

Separating real and imaginary parts, we obtain

$$p = e^{-2t} \cos 2u \tag{10}$$

$$q = -e^{-2t} \sin 2u \tag{11}$$

Squaring these two equations and adding gives the equation of the constant-t circles.

$$p^2 + q^2 = e^{-4t} \tag{12}$$

These circles are centered at the origin and have radii e^{-2t}

Dividing Eqs. (10) and (11) and inverting, we obtain

$$\frac{q}{p} = -\tan 2u \tag{13}$$

Substitution of $u = 2\pi w$ into Eq. (13) gives

$$\frac{q}{p} = -\tan 4\pi w \tag{14}$$

Hence the constant-w loci are straight lines passing through the origin and having a slope of $-\tan 4\pi w$.

PROBLEMS

1. Derive the equations for the constant-t and constant-u circles in the $r + jx$ plane of the rectangular impedance diagram. Show that these are circles and give the location of the centers and values of the radii.

2. A lossless coaxial line, having a characteristic impedance of 50 ohms, is $\frac{3}{8}$ of a wavelength long and is terminated in an impedance $Z_R = 75 + j60$ ohms. A condenser, having a capacitance of 4×10^{-12} farads is connected in series with the center conductor one-eighth of a wavelength from the receiving end. Using the impedance diagram, find the input impedance of the line at a frequency of 1,000 megacycles.

3. A line having a characteristic impedance of 50 ohms is used in the grid-plate circuit of an oscillator. The grid-plate capacitance is 1.8×10^{-12} farads and the line is tuned by means of a small air condenser at the far end. If the line is 15 centimeters long what value of capacitance would be required if the oscillator is to have a frequency of 600 megacycles?

4. A generator is connected to a load impedance by means of two coaxial lines in cascade. The first line is 0.6λ long and has $Z_{01} = 50$ ohms and $\alpha = 0.4$ neper per meter. The second line is 0.4λ long and has $Z_{02} = 75$ ohms and $\alpha = 0.3$ neper per meter. The load impedance is $Z_R = 45 - j75$ and the wavelength is $\lambda = 1$ meter. What is the input impedance?

5. A generator having an internal impedance Z_g is connected to a lossless transmission line having a characteristic impedance Z_0. Show that if $Z_g = Z_0$, the voltage at the receiving end of the open-circuited line will be equal to one-half the internal voltage of the generator, regardless of the length of line.

CHAPTER 10

TRANSMISSION-LINE NETWORKS

Transmission lines are often used as network elements in microwave systems. They may be used as resonant or antiresonant circuits, as reactive circuit elements in filter networks, as impedance transformers, as attenuators, or as circuit elements in various types of measuring systems.

Lumped-parameter networks are usually unsatisfactory at microwave frequencies, since the values of inductance and capacitance used in such networks are extremely small and slight variations due to mechanical vibration, temperature effects, etc., seriously alter the characteristics of the network. Transmission-line networks offer the advantages of greater stability, ease of adjustment, and much higher Q's than are possible with lumped-parameter circuits. The basic principles of transmission-line networks will be considered in this chapter.

10.01. Resonant and Antiresonant Lines.—The properties of lossless lines were considered in Sec. 8.06. Let us now see what effect losses have upon the input impedances of open-circuited and short-circuited lines.

The input impedance of short-circuited and open-circuited lines of length l are obtained by substituting l for x in Eqs. (8.07-3) and (8.08-3), respectively, yielding

$$Z = Z_0 \tanh \gamma l \quad \text{short-circuited line} \quad \text{(8.07-3)}$$

$$Z = Z_0 \coth \gamma l \quad \text{open-circuited line} \quad \text{(8.08-3)}$$

Now substitute $\gamma = \alpha + j\beta$ into these equations and use the identities

$$\tanh (\alpha + j\beta)l = \frac{\tanh \alpha l + j \tan \beta l}{1 + j \tanh \alpha l \tan \beta l} \quad \text{and} \quad \coth \gamma l = \frac{1}{\tanh \gamma l} \quad \text{to obtain}$$

$$Z = Z_0 \left(\frac{\tanh \alpha l + j \tan \beta l}{1 + j \tanh \alpha l \tan \beta l} \right) \quad \text{short-circuited line} \quad \text{(1)}$$

$$Z = Z_0 \left(\frac{1 + j \tanh \alpha l \tan \beta l}{\tanh \alpha l + j \tan \beta l} \right) \quad \text{open-circuited line} \quad \text{(2)}$$

For resonance or antiresonance, we have $\beta l = n\pi/2$, where n is an odd or even integer as specified at the end of Sec. 8.06. Consider the short-circuited line. Resonance occurs when n is even and antiresonance when n is odd. Inserting the corresponding values of βl into Eq. (1), together

176

with the approximation $\tanh \alpha l \approx \alpha l$ for small values of αl, we obtain the resonant and antiresonant impedances

$$Z = Z_0 \tanh \alpha l \approx Z_0 \alpha l \qquad \text{resonance} \qquad (3)$$

$$Z = Z_0 \coth \alpha l \approx \frac{Z_0}{\alpha l} \qquad \text{antiresonance} \qquad (4)$$

The resonant and antiresonant impedances of open-circuited lines are also given by Eqs. (3) and (4), respectively.

Let us now investigate the variation of impedance resulting from small frequency deviations either side of the resonant or antiresonant frequency. Let $\beta l = (n\pi/2) + \delta$, where δ is a small angular departure from the resonant or antiresonant value of βl. We then have

$$\tan \beta l = \frac{\tan (n\pi/2) + \tan \delta}{1 - \tan (n\pi/2) \tan \delta}$$

If n is even, this becomes $\tan \beta l = \tan \delta \approx \delta$, and if n is odd, $\tan \beta l = -(1/\tan \delta) \approx -(1/\delta)$. For the short-circuited line in the vicinity of resonance, n is even. Insertion of $\tan \beta l = \delta$ and $\tanh \alpha l = \alpha l$ into Eq. (1), gives

$$Z = Z_0 \left(\frac{\alpha l + j\delta}{1 + j\delta\alpha l} \right) \qquad \text{resonance} \qquad (5)$$

Similarly, for the short-circuited line in the vicinity of antiresonance, we have n odd, $\tan \beta l = -1/\delta$, and $\tanh \alpha l = \alpha l$, yielding

$$Z = Z_0 \left(\frac{1 + j\delta\alpha l}{\alpha l + j\delta} \right) \qquad \text{antiresonance} \qquad (6)$$

Equations (5) and (6) apply equally well for the open-circuited line in the vicinity of resonance and antiresonance. For small values of αl and δ, we have $\delta\alpha l \ll 1$, and the scalar values of the input impedance become approximately

$$Z = Z_0 \sqrt{(\alpha l)^2 + \delta^2} \qquad \text{resonance} \qquad (7)$$

$$Z = \frac{Z_0}{\sqrt{(\alpha l)^2 + \delta^2}} \qquad \text{antiresonance} \qquad (8)$$

It now remains to relate δ to the frequency. Let ω be the impressed angular frequency and ω_0 be the resonant or antiresonant angular frequency. From our previous assumption, we have $\beta l = \omega l/v = (n\pi/2) + \delta$. At the resonant or antiresonant frequency, we have $\omega_0 l/v = n\pi/2$. Combining these two expressions, we obtain

$$\delta = (\omega - \omega_0) \frac{l}{v} \qquad (9)$$

The quantity $\omega - \omega_0$ represents the difference between the impressed angular frequency and the resonant or antiresonant angular frequency.

10.02. The Q of Resonant and Antiresonant Lines.—One of the principal advantages of transmission-line networks is the high values of Q which are attainable by this means. Whereas a Q of 300 represents a relatively high value for lumped L-C circuits, Q's of the order of several thousand are attainable with lines. A high Q implies a high degree of frequency selectivity and therefore a sharply tuned circuit.

The Q of lumped L-C circuits is usually defined by the ratio of reactance to resistance, *i.e.*,

$$Q = \frac{\omega L}{R} \tag{1}$$

where ωL is the inductive reactance and R is the circuit resistance.

A more general definition of Q, which is applicable to any case of electrical or mechanical resonance, is

$$Q = 2\pi \frac{\text{peak energy storage}}{\text{energy dissipated per cycle}} \tag{2}$$

Multiplying the numerator and denominator of Eq. (2) by the frequency, and remembering that the product of energy dissipation per cycle times frequency is the power loss, we have

$$Q = \omega \frac{\text{peak energy storage}}{\text{average power loss}} = \frac{\omega W_S}{P_L} \tag{3}$$

where W_S is the peak energy storage and P_L is the time-average power loss.

In order to show that Eqs. (1) and (3) are equivalent, multiply the numerator and denominator of Eq. (1) by $\frac{1}{2}I^2$, where I is the amplitude of the current flowing in the inductance. This gives

$$Q = \omega \frac{\frac{1}{2}LI^2}{\frac{1}{2}I^2R} = \omega \frac{W_S}{P_L} \tag{4}$$

where $W_S = \frac{1}{2}LI^2$ is the peak energy storage in the inductance and $P_L = \frac{1}{2}I^2R$ is the time-average power loss.

Let us now apply the definition given by Eq. (3) to derive an expression for the Q of resonant and antiresonant lines. The final expression for Q is the same regardless of whether we choose an open-circuited or a short-circuited line operating at resonance or antiresonance.

Consider an antiresonant short-circuited line. The peak energy storage may be evaluated without serious error by assuming that the line is lossless. The voltage and current at any point on the line are then in time quadrature; therefore the energy storage in the electric field has its maxi-

mum value when the energy storage in the magnetic field is zero, and vice versa. The peak values of the energy storage in the electric and magnetic fields are equal and we may choose either for the purpose of evaluating the Q.

Consider the peak energy storage in the electric field. The capacitance of a differential length of line dx is $C\,dx$ and the peak energy storage is $dW_S = \frac{1}{2}(C\,dx)V^2$, where V is the voltage amplitude. Substituting the voltage from Eq. (8.06-5) into this expression and integrating between the limits $\beta x = 0$ and $\beta x = n\pi/2$ (where n is odd for the antiresonant line), the peak energy storage becomes

$$W_S = \frac{C(I_R Z_0)^2}{2\beta} \int_0^{n\pi/2} \sin^2\beta x \, d(\beta x) = \frac{n\pi C}{8\beta}(I_R Z_0)^2 \tag{5}$$

where I_R is the amplitude of the current at the receiving end.

The time-average power loss is obtained from $P_L = \frac{1}{2}VI\cos\theta$, where V and I are the amplitudes of the voltage and current at the input terminals, and θ is the phase angle between V and I. The voltage and current at the input terminals are found by inserting $\beta l = n\pi/2$, $\cosh\alpha l \approx 1$, and $\sinh\alpha l \approx \alpha l$ into Eqs. (8.07-4) and (8.07-5), yielding the scalar values $V = I_R Z_0$ and $I = I_R \alpha l$, and $\theta = 0$. The time-average power loss is therefore

$$P_L = \frac{I_R^2 Z_0 \alpha l}{2} \tag{6}$$

Inserting Eqs. (5) and (6) into (3), together with $\beta l = n\pi/2$, $Z_0 = \sqrt{L/C}$, and $\beta = \omega\sqrt{LC}$, we obtain an expression for the Q

$$Q = \frac{\beta}{2\alpha} \tag{7}$$

Thus, the Q is the phase constant divided by twice the attenuation constant. It is also interesting to note that if we neglect the conductance in Eq. (8.04-6) so that $\alpha = R/2Z_0$, and insert this, together with the above expressions for β and Z_0, into Eq. (7), the expression for the Q reduces to the familiar form $Q = \omega L/R$ where L and R are the series inductance and resistance per unit length of line.

The Q is a measure of the frequency selectivity of a resonant or antiresonant circuit. To show this relationship, we return to Eqs. (10.01-7 and 8). Let ω represent the angular frequency for which $\delta = \alpha l$. At this frequency, the input impedance is $\sqrt{2}$ times the resonant impedance, or $1/\sqrt{2}$ times the antiresonant impedance, as the case may be. Substitution of δ from Eq. (10.01-9) into the above expression yields $(\omega - \omega_0)/v_c = \alpha$, where ω_0 is the resonant angular frequency. From Eq. (7) we obtain

$Q = \beta/2\alpha = \omega_0/2\alpha v_c$; hence

$$Q = \frac{\omega_0}{2(\omega - \omega_0)} = \frac{f_0}{2\Delta f} \tag{8}$$

where f_0 is the resonant frequency and $\Delta f = (\omega - \omega_0)/2\pi$.

The resonant and antiresonant impedances may be expressed in terms of the Q, by inserting $\beta l = n\pi/2$ and α from Eq. (7) into (10.01-3) and (10.01-4), yielding

$$Z = \frac{Z_0 n\pi}{4Q} \qquad \text{resonance} \tag{9}$$

$$Z = \frac{4Z_0 Q}{n\pi} \qquad \text{antiresonance} \tag{10}$$

In general, the Q of a line is increased by increasing either the size of the conductors or the spacing between conductors. Increasing the size of the conductors decreases the skin-effect resistance, whereas increasing the spacing between conductors increases the inductance per unit length of line. However, in open-wire lines, the radiation losses increase as the separation distance increases.

For the coaxial line, the attenuation constant, as given in Table 1, becomes a minimum when the ratio b/a is 3.6. This corresponds to a characteristic impedance of approximately 77 ohms. Since β is independent of b and a, it follows from Eq. (7) that maximum Q likewise occurs when $b/a = 3.6$. Figure 1 is a plot of the Q of copper coaxial lines as a function of frequency for various sizes of lines, all having the optimum value $b/a = 3.6$.

FIG. 1.—Q of copper coaxial lines having the optimum ratio $b/a = 3.6$.

10.03. Lines with Reactance Termination.—Lines are often used as tuned elements in vacuum-tube circuits where they are shunted by the interelectrode capacitances and conductances of the tube. A line may be tuned by means of a small variable condenser shunted across either end of the line. Let us observe what effect lumped reactances have upon the resonant and antiresonant frequencies.

In high-Q circuits the resistance has very little effect upon the resonant frequency; hence we shall confine our attention to lossless lines which are terminated by pure reactances. Consider the lines shown in Fig. 2, with lumped reactances X_1 at the sending end and X_R at the receiving end. In Fig. 2a the reactance X_1 is assumed to be connected in series with a zero-impedance generator. This is equivalent to the series L-C circuit of Fig. 2b, the input impedance having zero value at the resonant frequency. In Fig. 2c, the reactance X_1 is assumed to be shunted across an infinite-impedance generator, this being analogous to the parallel L-C circuit of Fig. 2d.
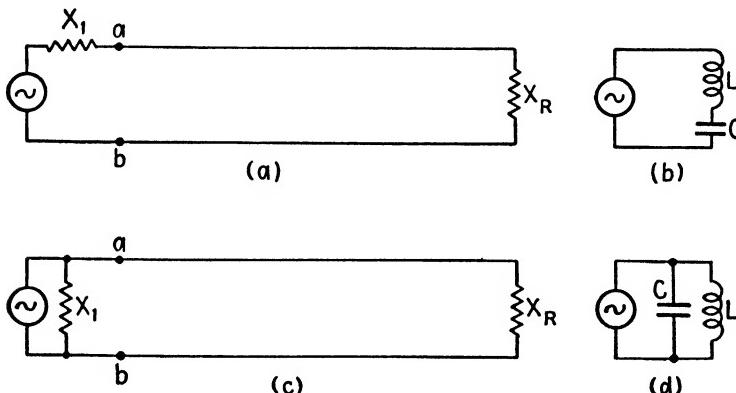
FIG. 2.—Lines with reactance terminations and their lumped-circuit equivalents.

A condition of resonance exists when the reactances looking both ways at any pair of terminals, such as at ab in Fig. 2a or 2c, are equal in magnitude and opposite in sign. In Fig. 2a this results in a resonant input impedance, whereas in Fig. 2b, the input impedance is antiresonant.

The input reactance of the line alone at terminals ab is obtained from Eq. (8.06-4),

$$Z = jZ_0 \left(\frac{X_R + Z_0 \tan \beta l}{Z_0 - X_R \tan \beta l} \right) \tag{1}$$

Applying the above criterion, resonance or antiresonance occurs when

$$X_1 = -Z_0 \left(\frac{X_R + Z_0 \tan \beta l}{Z_0 - X_R \tan \beta l} \right) \tag{2}$$

Solving for $\tan \beta l$ and substituting $\beta = \omega/v_c$, we have

$$\tan \frac{\omega l}{v_c} = Z_0 \left(\frac{X_1 + X_R}{X_1 X_R - Z_0^2} \right) \tag{3}$$

In these equations X_1 and X_R take positive values for inductive reactance and negative values for capacitive reactance.

If the line is short-circuited at the receiving end, we have $X_R = 0$, and Eq. (3) reduces to

$$\tan \frac{\omega l}{v_c} = -\frac{X_1}{Z_0} \tag{4}$$

and if open-circuited, $X_R = \infty$, and

$$\tan \frac{\omega l}{v_c} = \frac{Z_0}{X_1} \tag{5}$$

If there is no reactance at the sending end, we set $X_1 = 0$ for the series connection and $X_1 = \infty$ for the parallel connection.

Resonant lines have a multiplicity of resonant and antiresonant frequencies which may be either harmonically or inharmonically related. If the line contains no lumped reactance, the resonant and antiresonant frequencies are harmonically related. For example, the short-circuited line without reactance termination at either end is resonant when $\tan \omega l/v_c = 0$ and antiresonant when $\tan \omega l/v_c = \pm \infty$. The corresponding resonant and antiresonant frequencies are given by

$$f = \frac{n v_c}{4l} \quad \text{or} \quad l = \frac{n\lambda}{4} \tag{6}$$

where n is an even integer for resonance and odd for antiresonance. The resonant and antiresonant frequencies are therefore harmonically related.

In general, if the line is terminated by lumped reactances the resonant and antiresonant frequencies are not harmonically related. This is apparent since the frequency appears on both sides of Eq. (3). The question sometimes arises: how can we design a line having reactance terminations so as to be simultaneously resonant or antiresonant at any two or more specified frequencies? Such a problem might arise if we were to design an oscillator or amplifier to deliver a relatively large output at a harmonic of the fundamental frequency. The interelectrode capacitance of the tube constitutes a lumped reactance shunting the input end of the line, as shown in Fig. 2. Therefore the antiresonant frequencies will, in general, be inharmonically related unless we specifically design the circuit to have harmonic antiresonant frequencies.

Equation (10.03-3) shows that there are four variables which we are at liberty to adjust. These are the two terminating reactances, the characteristic impedance of the line, and the length of line. By proper adjustment of the four variables, it should be possible to make the circuit either resonant or antiresonant at four specified frequencies which may be either harmonically or inharmonically related.

As a specific example, consider a short-circuited line with a capacitance C shunted across its input terminals. We wish to find the length of line

and value of C required to make the circuit antiresonant at the two angular frequencies ω_1 and ω_2. For the two specified frequencies Eq. (4) becomes

$$\tan \frac{\omega_1 l}{v_c} = \frac{1}{\omega_1 C Z_0} \qquad \tan \frac{\omega_2 l}{v_c} = \frac{1}{\omega_2 C Z_0} \tag{7}$$

Dividing these equations gives

$$\frac{\tan \omega_1 l / v_c}{\tan \omega_2 l / v_c} = \frac{\omega_2}{\omega_1} \tag{8}$$
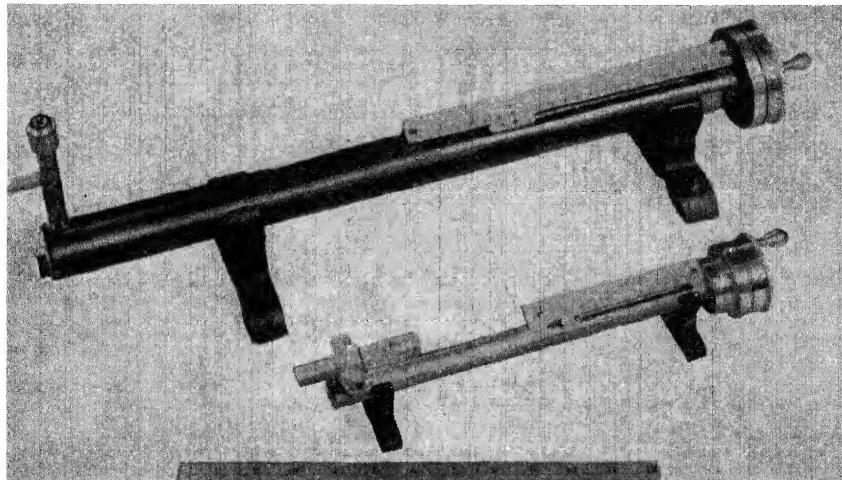
FIG. 3.—Coaxial wavemeters.

If the two frequencies are harmonically related, we have $\omega_2 = n\omega_1$, and Eq. (8) reduces to

$$\tan \frac{\omega_1 l}{v_c} = n \tan n \frac{\omega_1 l}{v_c} \tag{9}$$

This is a transcendental equation of the form $\tan \theta = n \tan n\theta$. Solutions for $\omega_1 l / v_c$ may be obtained either by Newton's method [1] or by graphical methods. The length of line may then be computed from the known value of $\omega l_1 / v_c$.

10.04. Measurement of Wavelength.—The coaxial wavemeter shown in Fig. 3 consists of a coaxial line which is short-circuited at both ends. One end contains a sliding piston which is adjustable by means of a worm gear. The microwave source is coupled to the wavemeter by means of a coaxial line which terminates in a small coupling loop inside the coaxial

---

[1] DOHERTY, R. E., and E. G. KELLER, "Mathematics in Modern Engineering," chap. 4, John Wiley & Sons, Inc., New York, 1936.

wavemeter. A second coupling loop is connected to a crystal detector and microammeter for a resonance indicator. Resonance occurs when the co-axial line is an integral number of half wavelengths long. The micro-ammeter reading has its maximum value at resonance and decreases abruptly as the coaxial wavemeter is tuned away from resonance. A centimeter scale and vernier on the dial indicate the wavelength.

10.05. Measurement of Impedances at Microwave Frequencies.—An unknown impedance can be measured by connecting it to a low-loss transmission line having a known value of characteristic impedance and measuring: (1) the standing wave ratio $\rho = |V_{max}|/|V_{min}|$ and (2) the distance from the terminating impedance to the first voltage maximum or minimum.

The standing wave ratio is a measure of the impedance mismatch of a line. If the line is terminated in its characteristic impedance, the standing wave ratio is unity. The standing wave ratio increases as the terminal impedance becomes increasingly greater than or less than the character-istic impedance, approaching infinite value for the open-circuited or short-circuited line.

For a lossless line, the voltage maximum and current minimum occur at the same point on the line. In Sec. 9.05 it was shown that the voltage maximum may be represented by $|V_{max}| = V'_R(1 + |r_R|)$, where $|r_R|$ is the magnitude of the reflection coefficient. By a similar method, the current minimum can be shown to be $|I_{min}| = (1 - |r_R|)(V'_R/Z_0)$. At the point where the voltage is a maximum and the current a minimum, the impedance is

$$Z = \frac{|V_{max}|}{|I_{min}|} = Z_0 \frac{1 + |r_R|}{1 - |r_R|} = \rho Z_0 \tag{1}$$

where ρ is the standing-wave ratio given by Eq. (9.05-4). In a similar manner, it may be shown that the impedance at the voltage minimum (current maximum) is

$$Z = \frac{Z_0}{\rho} \tag{2}$$

Now write Eq. (8.06-4) for the impedance at the point of maximum voltage on the line, letting $x = x_{max}$ and $Z = \rho Z_0$,

$$\rho Z_0 = Z_0 \left(\frac{Z_R + jZ_0 \tan \beta x_{max}}{Z_0 + jZ_R \tan \beta x_{max}} \right) \tag{3}$$

Solving this for Z_R and substituting $\beta = 2\pi/\lambda$, we have

$$Z_R = Z_0 \left(\frac{\rho - j \tan \dfrac{2\pi x_{max}}{\lambda}}{1 - j\rho \tan \dfrac{2\pi x_{max}}{\lambda}} \right) \tag{4}$$

If ρ, Z_0, x_{max}, and λ are known, the terminal impedance may be readily computed from Eq. (4). A similar relationship may be derived for the terminating impedance in terms of the distance from x_{min} to the voltage minimum.

The impedance diagram can be used to evaluate an unknown load impedance. As an example, assume that the values $\rho = 2$ and $x_{max}/\lambda = 0.15$
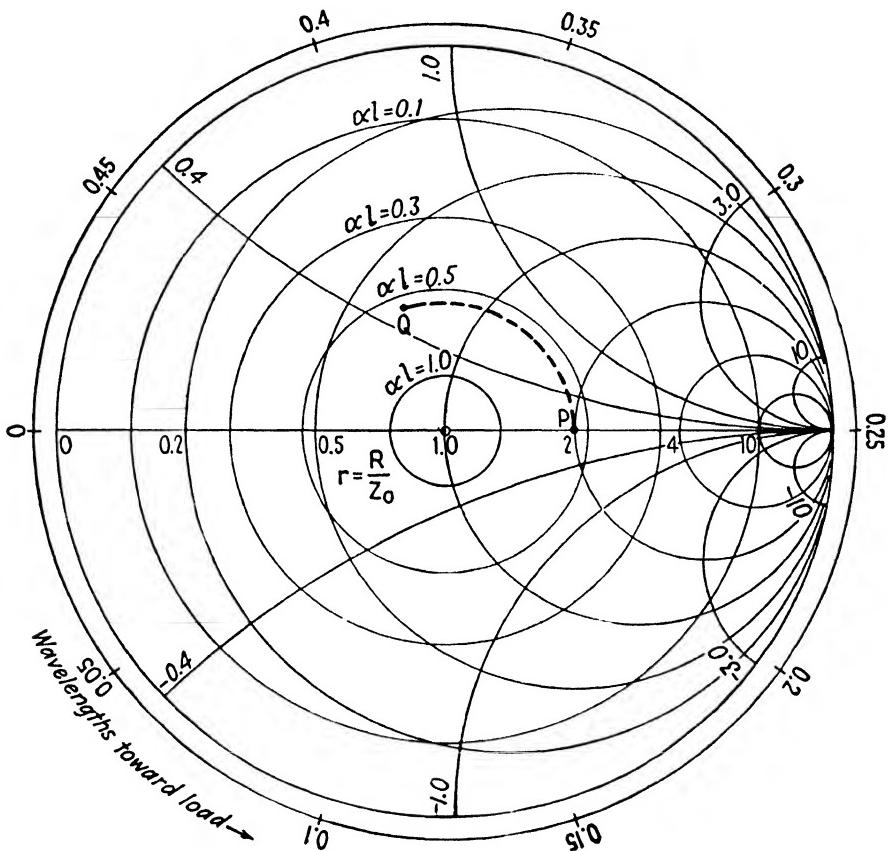
FIG. 4.—Use of the impedance diagram for determining an unknown impedance.

have been experimentally determined. The normalized impedance at the voltage-maximum point on the line, from Eq. (1), is $Z/Z_0 = \rho = 2.0 + j0$. Enter the impedance diagram at this value of impedance (point P in Fig. 4) and observe the corresponding values of $\alpha l = 0.54$ and $w_0 = 0.25$. Now proceed in a counterclockwise direction on the constant-αl circle to point Q where $w = w_0 + (x_{max}/\lambda) = 0.40$ (on the "wavelengths toward the load" scale). This corresponds to moving from the voltage-maximum point to

the receiving end of the line.   The normalized impedance at point Q is $Z_R/Z_0 = 0.68 + j0.48$.   This is the normalized load impedance.

Figure 5a shows a standing-wave indicator which is used to measure the standing-wave ratio at wavelengths of from approximately 5 to 15 centi-
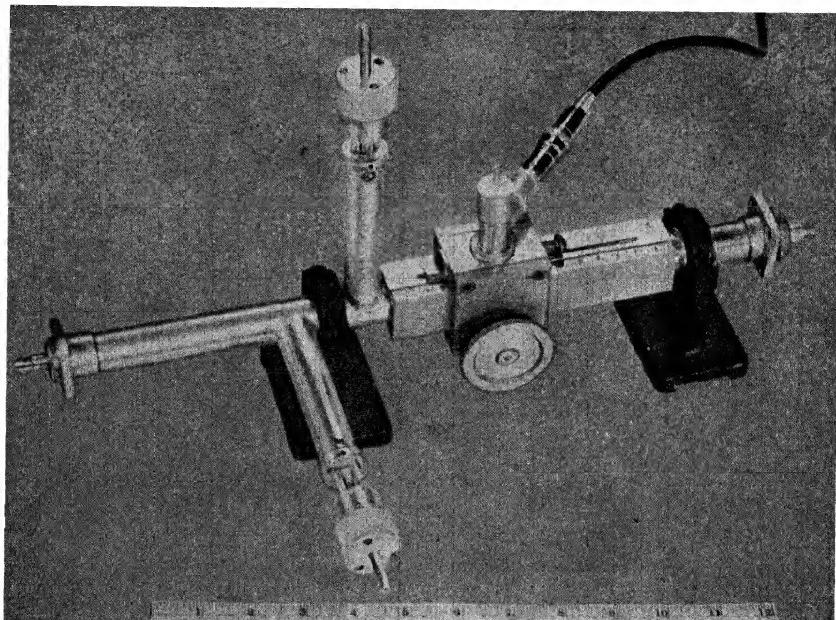
(a)
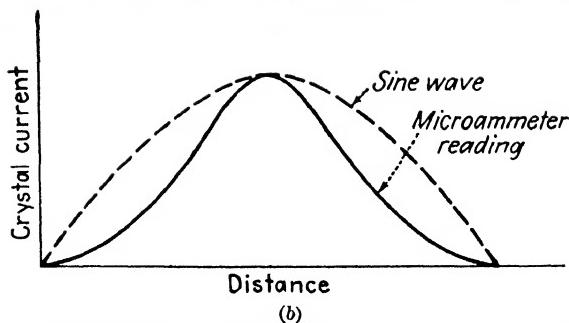
(b)

Fig. 5.—Standing-wave indicator and method of crystal calibration.

meters.   This consists of a sliding probe which extends a short distance into a slotted section of coaxial line.   The probe is connected to a crystal detector and microammeter.   This device actually measures the electric intensity in the coaxial line.   However, the electric intensity is proportional to the voltage between conductors, hence the standing-wave indi-

cator may be assumed to be a voltage-measuring device. A scale on the standing-wave indicator is used to measure the distance between the terminal impedance and the voltage maximum or minimum.

To calibrate the standing-wave indicator, the coaxial line is short-circuited at its output terminals and data are taken for a curve of micro-ammeter reading plotted against probe position, as shown in Fig. 5b. The sine-wave curve, also shown in Fig. 5b, is the curve which would be obtained if the crystal had a linear volt-ampere characteristic. The relative calibration curve is then obtained by plotting a curve representing the sine wave values as ordinates and the corresponding microammeter readings as abscissas. Care must be taken to couple the probe loosely to the coaxial line in order to avoid disturbing the standing wave on the coaxial line.
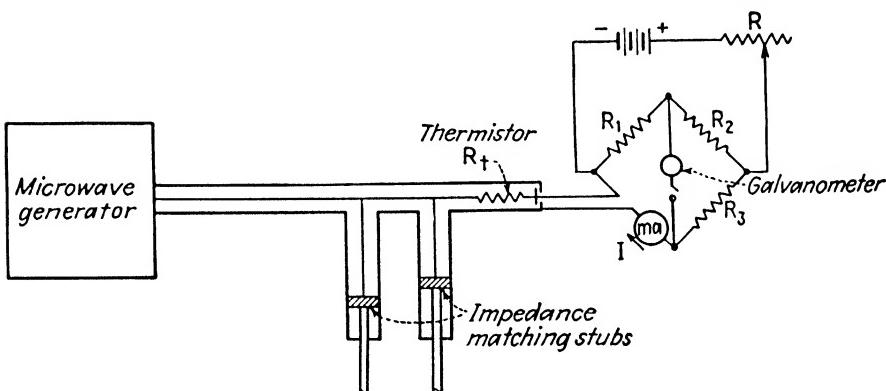
FIG. 6.—Thermistor bridge for power measurement.

10.06. Power Measurement at Microwave Frequencies.—It is often necessary to measure the power output of oscillators or amplifiers operating at microwave frequencies. The thermistor bridge or bolometer bridge, shown in Fig. 6, provides a convenient and reasonably accurate method of measuring power. One arm of the bridge, designated as R_t in Fig. 6, contains a resistance element which has a relatively high temperature coefficient of resistance. This element is connected in such a manner as to absorb the power from the microwave source whose power output is being measured.

With the microwave source disconnected, the bridge is balanced and the milliammeter reading in the R_t branch is observed. The microwave source is then connected and the stubs are adjusted for maximum power transfer to R_t, as indicated by a maximum unbalance of the bridge. The unbalance results from the fact that the value of R_t changes as the power dissipation in it increases. Balance is restored by decreasing the d-c power loss in R_t by an amount exactly equal to the microwave power dissipation in this resistance. This is accomplished by means of the rheostat in the

battery circuit. If I_1 is the direct current through R_t for the initial balance and I_2 is the direct current when the bridge is balanced with the microwave source connected, the microwave power is $P_{ac} = (I_1^2 - I_2^2)R_t$. The milliammeter may be calibrated to read the power in watts.

If the element R_t is a semiconductor having a negative temperature coefficient of resistance, it is known as a thermistor. Negative temperature-coefficient materials suitable for thermistors include uranium oxide, a mixture of nickel oxide and manganese oxide, and silver sulphide. Conducting wires are sometimes used which have a positive temperature coefficient of resistance. These, are known as bolometers or barretters. One commercial form consists of a straight platinum wire, 0.06 mil in diameter,
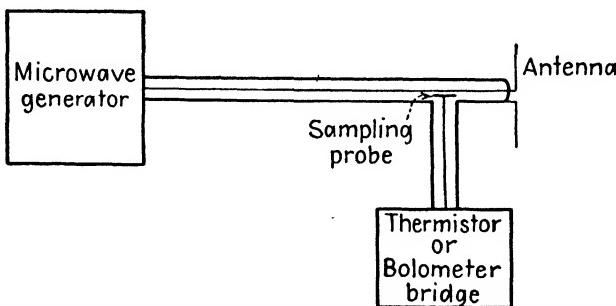
FIG. 7.—Power measurement by the sampling method.

which is mounted in a cylinder for use in standard coaxial line fittings. This particular element has a maximum power rating of 32.5 milliwatts and may be used to measure values of power as low as 10 microwatts.

In order to minimize skin-effect errors, thermistors and barretters usually have very small diameters; consequently their power rating is quite small. Higher power levels may be measured by inserting a calibrated attenuator between the source and R_t in Fig. 6. Sections of coaxial line having a high-loss dielectric are sometimes used as attenuators. However, the attenuation characteristics of this type of attenuator are likely to vary with temperature and humidity. Another type of attenuator consists of a glass rod upon which has been deposited a thin layer of platinum. The platinized glass rod is used as the center conductor in a coaxial line which is inserted between the power source and the thermistor or bolometer bridge.

The power consumed in a load impedance may be measured by a sampling method shown in Fig. 7. In this method the power-measuring circuit is loosely coupled to the coaxial line which connects the microwave source to the load. A small fraction of the total power enters the power-measuring bridge, which may be of the thermistor or bolometer type. In this method, it is necessary to calibrate the power-measuring circuit with the given load impedance in order to determine what fraction of the total power is being

drawn off by the power-measuring circuit. A more accurate sampling method consists of a coaxial line containing a directional coupler, similar to that shown for wave guides in Chap. 18.

10.07. Effect of Impedance Mismatch upon Power Transfer.—A well-known power-transfer theorem states that if a variable load impedance is connected to a constant-voltage generator, maximum power transfer occurs when the load impedance is equal to the complex conjugate of the generator impedance. When this condition prevails the impedances are said to be matched. It is sometimes helpful to be able to determine the power sacrificed as a result of not having matched impedances.

FIG. 8.—Generator and load.

Referring to Fig. 8, let $Z_g = R_g + jX_g$ be the generator impedance and $Z_L = R_L + jX_L$ represent the load impedance. The scalar value of current is

$$|I| = \frac{V_g}{\sqrt{(R_g + R_L)^2 + (X_g + X_L)^2}} \tag{1}$$

and the power consumed in the load is

$$P = |I|^2 R_L = \frac{V_g^2 R_L}{(R_g + R_L)^2 + (X_g + X_L)^2} \tag{2}$$

If the load impedance is the only variable, the power is a maximum when $R_L = R_g$ and $X_L = -X_g$, that is, when the generator and load impedance are conjugates. The power is then

$$P_{\max} = \frac{V_g^2}{4R_g} \tag{3}$$

The ratio of the power for any load impedance to the maximum power is found by dividing Eq. (2) by (3),

$$\frac{P}{P_{\max}} = \frac{4R_g R_L}{(R_g + R_L)^2 + (X_g + X_L)^2} \tag{4}$$

Now divide the numerator and denominator by R_g^2 and let $R' = R_L/R_g$ and $X' = (X_g + X_L)/R_g$, giving

$$\frac{P}{P_{\max}} = \frac{4R'}{(1 + R')^2 + (X')^2} \tag{5}$$
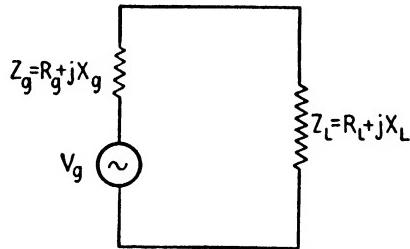
If we let R' and X' be the coordinate axes in a rectangular coordinate system, the loci representing constant values of P/P_{max} are circles as shown in Fig. 9. This graph makes it possible to evaluate the ratio of actual power to maximum power for any condition of impedance mismatch. The procedure is to compute R' and X' for the particular problem. The value
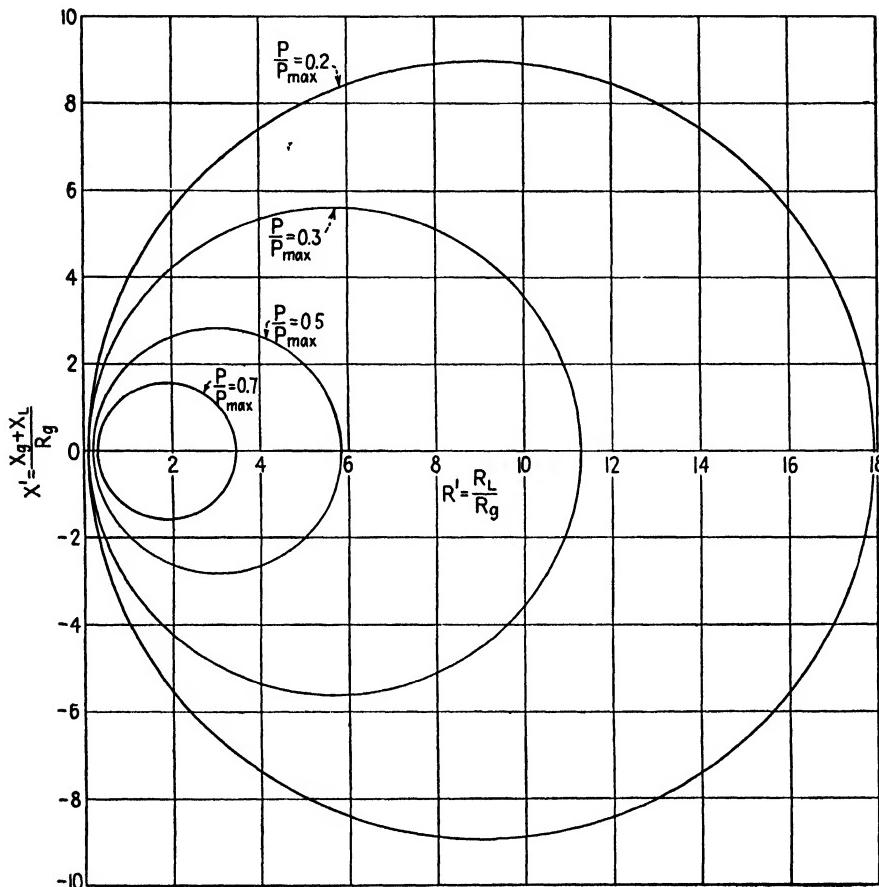
FIG. 9.—Power transfer ratio P/P_{max} for mismatched impedances.

of P/P_{max} is then obtained from Fig. 9. If the load consists of a transmission line with a load impedance at the distant end, the impedance Z_L is the input impedance of the line.

10.08. Power-transfer Theorem.—There is an interesting and useful power-transfer theorem which states that if a generator is connected through one or more pure reactance networks to a load, as shown in Fig. 10, and the conditions are such that there is a conjugate impedance match at one pair of terminals, then there will be a conjugate impedance match

at every pair of terminals and maximum power will be transferred from the generator to the load.

This may be readily verified by the use of Thévenin's theorem. If we break into the network at any junction, such as at ab in Fig. 10a, Thévenin's theorem permits us to replace the network to the left of ab by a generator as shown in Fig. 10b. The generator impedance Z'_g in the equivalent network is equal to the impedance looking to the left at terminals ab, and the generator voltage V'_g is the open-circuit voltage at ab. The network to the right of ab in Fig. 10a is replaced by its equivalent impedance Z_2 in Fig. 10b. Now assume that there is a conjugate match of impedances
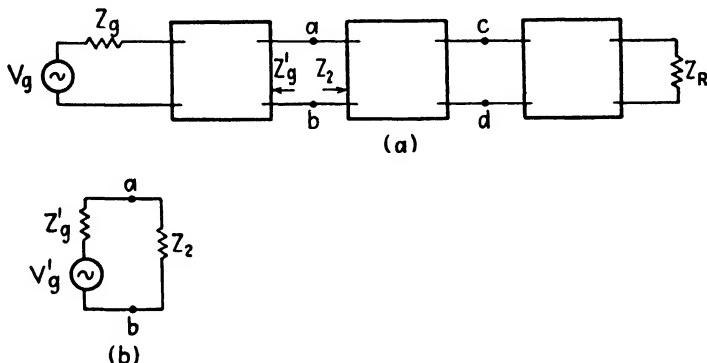
FIG. 10.—(a) Network consisting of a generator connected to a load impedance through pure reactance networks, and (b) equivalent circuit from Thévenin's theorem.

at ab in Figs. 10a and 10b. A conjugate impedance match in the circuit of Fig. 10b signifies that there is a maximum power transfer past the junction ab. If there is a maximum power transfer in the equivalent circuit, there must likewise be maximum power transfer in the original circuit. Since we have maximum power flow past the junction ab in Fig. 10a, and there is no power lost in the reactive networks, it follows that there is a maximum power flow at every junction and likewise maximum power transfer to the load. Consequently, there must be a conjugate impedance match at every junction, since, if there were not a conjugate impedance match at any junction, there could not be maximum power flow past this junction.

For most transmission lines operating at microwave frequencies, we have $\omega L \gg R$ and $\omega C \gg G$; hence the lines may be treated as pure reactance networks. The foregoing power-transfer theorem makes it possible to match impedances at any point on the line between the generator and load and be assured of a conjugate impedance match throughout the entire system, resulting in maximum power transfer to the load.

10.09. Quarter-wavelength and Half-wavelength Lines.—Transmission lines which are either a quarter wavelength or a half wavelength long have

special impedance-transforming properties. Consider the input impedance to a quarter-wavelength lossless line which is terminated in an impedance Z_R. Inserting $\beta l = \pi/2$ into Eq. (8.06-4), we obtain

$$Z = \frac{Z_0^2}{Z_R} \tag{1}$$

The input impedance varies inversely as Z_R; therefore the quarter-wavelength line is effectively an impedance inverter. If Z_R is inductive, the input impedance is capacitive, and vice versa. If the load impedance is constant, it is possible to control the magnitude of the input impedance,
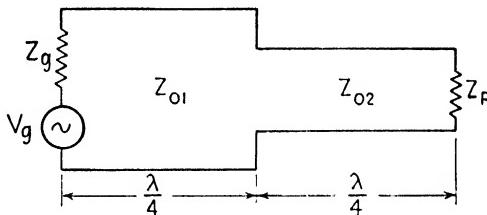
Fig. 11.—Impedance transformer consisting of two quarter-wavelength lines.

but not its phase angle, by a proper choice of Z_0. If a generator having an impedance Z_g is connected to the input end of the line and both Z_g and Z_R are pure resistances, maximum power transfer occurs when

$$Z_g = \frac{Z_0^2}{Z_R} \quad \text{or} \quad Z_0 = \sqrt{Z_g Z_R} \tag{2}$$

From a practical point of view, the range of characteristic impedances of coaxial lines is from 5 to 250 ohms, whereas that for open-wire lines is from 90 to 1,000 ohms. These provide the practical limits of impedance transformation using the quarter-wavelength line.

Let us now see what effect variations in frequency have upon the power transfer. Assume that a lossless line, terminated by an impedance Z_R, has matched impedances at the generator at the frequency for which the line is a quarter wavelength long. If Z_g and Z_R are approximately equal, the variation in power transfer with frequency will be relatively small. However, if Z_g and Z_R differ greatly the power transfer decreases rapidly as the frequency departs from the frequency at which the line is a quarter wavelength long, and therefore the impedance transformer is highly frequency selective.

Two or more quarter-wavelength sections of line having different characteristic impedances, such as shown in Fig. 11, may be used to obtain an impedance transformer which is less frequency selective than that of a single quarter-wavelength section.

Now consider the input impedance of a half-wavelength line which is terminated in an impedance Z_R.   Inserting $\beta l = \pi$ into Eq. (8.06-4), we obtain

$$Z = Z_R \tag{3}$$

Therefore, the input impedance of the half-wavelength line is equal to the terminating impedance, or the half-wavelength line is effectively a one-to-one ratio transformer.

10.10. Single-stub Impedance Matching.—From a practical point of view, the use of the quarter-wavelength line as an impedance transformer is restricted largely to the matching of resistive impedances where the frequency and impedances are constant.   Stub impedance-matching systems
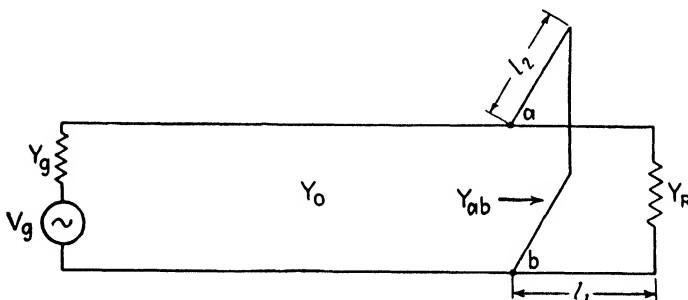
FIG. 12.—Single-stub impedance matching.

are more versatile in that they may be used to match complex impedances and are more readily adjustable.   A stub consists of an open-circuited or short-circuited line which is shunted across the transmission line between the generator and the load.   One or more such stubs may be used for impedance matching.

If a single stub is used, as shown in Fig. 12, it is necessary to have both the length of the stub and the distance from the stub to the load adjustable in order to match all possible load impedances.   Let us analyze this case. Since we are dealing with parallel circuits, it is convenient to use admittances.   The characteristic admittance of the line $Y_0 = 1/Z_0$ may be assumed to be a pure conductance for a low-loss line.

Maximum power transfer requires a conjugate admittance (or impedance) match at the generator and, likewise, a conjugate admittance match at points ab where the stub is connected to the transmission line.   In the following discussion, it is assumed that the line is lossless and therefore that the characteristic admittance Y_0 is a pure conductance.   If the generator admittance happens to be equal to the characteristic admittance, that is, if $Y_g = Y_0$, then the admittance looking to the left at ab is Y_0 and·the admittance looking to the right at ab (including the stub) must likewise

be equal to Y_0 for maximum power transfer. Under these conditions, there will be no standing waves on the line between the generator and the stub, although standing waves will exist on the section of line from the stub to the load and on the stub itself.

If the generator admittance is not equal to Y_0, then the admittance looking to the left at ab is not equal to Y_0, and therefore the admittance looking to the right at ab must likewise have some value other than Y_0 for a conjugate admittance match. The section of line between the generator and stub then serves as an impedance transformer and has standing waves of voltage and current along its length. It should be noted that the absence of standing waves on a line is not always an indication of maximum power transfer.

In the following analysis, we shall assume a lossless line with $Y_g = Y_0$. Maximum power transfer then requires that $Y_{ab} = Y_0$, where Y_{ab} is the admittance looking to the right at ab including the stub. The admittance of a lossless stub is a pure susceptance. The stub is located at that point on the line where the real part of the admittance, looking toward the load, is Y_0. The stub length is then adjusted so that its susceptance is equal and opposite to the susceptance of the line at this point, thereby canceling the susceptance and leaving $Y_{ab} = Y_0$.

The reflection coefficient equations will be used in order to gain facility in the use of these equations. The admittance at any point on the line looking toward the load (with the stub disconnected) is obtained by inverting Eq. (9.01-4)

$$\frac{Y}{Y_0} = \frac{1 - r_R e^{-2\gamma x}}{1 + r_R e^{-2\gamma x}} \tag{1}$$

where the reflection coefficient is given by $r_R = (Y_0 - Y_R)/(Y_0 + Y_R)$. Now let

$$r_R = \frac{Y_0 - Y_R}{Y_0 + Y_R} = |\, r_R \,| e^{-j2u_0} \tag{2}$$

where $|\, r_R \,|$ is the absolute value of the reflection coefficient and $-2u_0$ is the angle. Substitution of Eq. (2) into (1), together with $\gamma = j\beta$ (for a lossless line), gives

$$\frac{Y}{Y_0} = \frac{1 - |\, r_R \,| e^{-2j(u_0 + \beta x)}}{1 + |\, r_R \,| e^{-2j(u_0 + \beta x)}} = \frac{1 - |\, r_R \,| e_i^{-j\psi}}{1 + |\, r_R \,| e^{-j\psi}} \tag{3}$$

where $\psi = 2(u_0 + \beta x)$

Using the trigonometric identity $e^{-j\psi} = \cos\psi - j\sin\psi$ and separating the real and imaginary terms, we obtain from Eq. (3),

$$\frac{Y}{Y_0} = \frac{1 - |\, r_R \,|^2}{1 + 2|\, r_R \,|\cos\psi + |\, r_R \,|^2} + \frac{j2|\, r_R \,|\sin\psi}{1 + 2|\, r_R \,|\cos\psi + |\, r_R \,|^2} \tag{4}$$

The stub is located at that point on the line where the real part of the admittance is equal to Y_0, or where the real part of Eq. (4) has unit value. This requires that

$$\cos \psi = -| r_R | \tag{5}$$

Inserting $\psi = 2(u_0 + \beta l_1)$, where l_1 is the distance between the receiving end of the line and the stub, we have

$$\cos 2(u_0 + \beta l_1) = -| r_R |$$

$$l_1 = \frac{\lambda}{2\pi} \left(-u_0 + \frac{1}{2} \cos^{-1} - | r_R | \right) \tag{6}$$

The values of $| r_R |$ and u_0 may be evaluated in terms of known values of Y_0 and Y_R by using Eq. (2). These are then substituted into Eq. (6) to obtain the distance l_1 from the load to the stub.

From Eq. (5), we obtain $\sin \psi = \pm \sqrt{1 - | r_R |^2}$. Substitution of this, together with Eq. (5), into (4), yields

$$\frac{Y}{Y_0} = 1 + \frac{j2| r_R |}{\pm \sqrt{1 - | r_R |^2}} \tag{7}$$

For a conjugate admittance match, under the assumed conditions, the susceptance portion of Eq. (7) must be canceled by an equal and opposite stub susceptance. Assume that the characteristic admittance of the stub and line are equal. The normalized stub admittance must then be

$$\frac{Y_{\text{stub}}}{Y_0} = - \frac{j2| r_R |}{\pm \sqrt{1 - | r_R |^2}} \tag{8}$$

The normalized admittances of the short-circuited and open-circuited stubs are obtained from Eqs. (8.06-7 and 10),

$$\frac{Y_{\text{stub}}}{Y_0} = -j \cot \frac{2\pi l_2}{\lambda} \quad \text{short-circuited stub} \tag{9}$$

$$\frac{Y_{\text{stub}}}{Y_0} = j \tan \frac{2\pi l_2}{\lambda} \quad \text{open-circuited stub} \tag{10}$$

The length of stub is found by equating either Eq. (9) or (10) to Eq. (8) and solving for l_2.

The position of the stub and its length may also be determined from standing-wave measurements on the line. With the stub disconnected, the maximum and minimum voltages are observed and the distance from the load to the first voltage maximum is obtained. Equation (9.05-4) is then used to obtain the value of $| r_R |$ and Eq. (9.05-7) yields u_0. These values may then be substituted into Eqs. (6) and (8) and the values of l_1 and l_2 are then determined as outlined above.

Example. A generator having an internal impedance $Z_g = 75$ ohms is connected to a load impedance $Z_R = 250$ ohms by means of a transmission line having a characteristic impedance of 75 ohms. The wavelength of the impressed signal is $\lambda = 20$ cm. Find the position and length of a short-circuited stub which will yield maximum power transfer to the load.
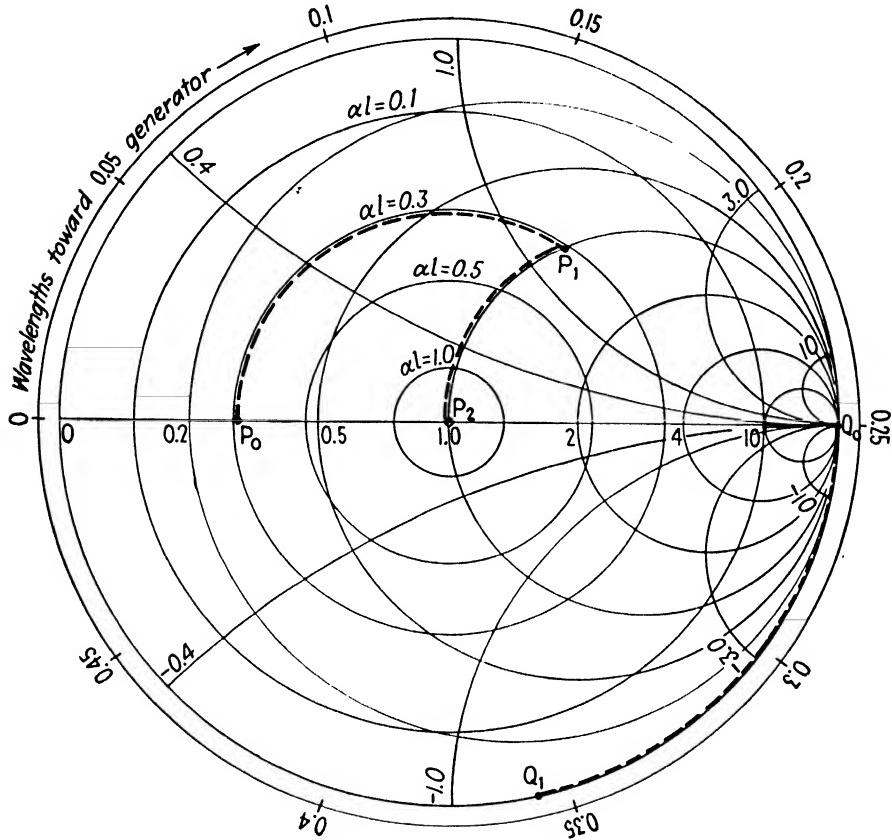
Fig. 13.—Graphical solution of single-stub impedance matching.

Consider first the analytical solution. The admittances are $Y_0 = 1/Z_0 = 0.0133$ mhos and $Y_R = 1/Z_R = 0.004$ mhos. Substitution of these values into Eq. (2) yields

$$r_R = |r_R| e^{-2ju_0} = 0.537$$

or $|r_R| = 0.537$ and $u_0 = 0$. Inserting these values into Eq. (6), we obtain

$$\frac{l_1}{\lambda} = \frac{1}{4\pi} \cos^{-1}(-0.537)$$

There are a multiplicity of values of l_1 which satisfy this equation. These values differ by a half wavelength and any one of the values may be used. The shortest distance

corresponds to the second-quadrant angle of 121.5° or 2.12 radians. This gives

$$\frac{l_1}{\lambda} = 0.169 \qquad l_1 = 3.38 \text{ cm.}$$

The length of the stub is found by equating (8) and (9) and substituting $|r_R| = 0.537$, yielding

$$\cot \frac{2\pi l_2}{\lambda} = 1.264$$

Again we find a multiplicity of solutions differing by a half wavelength. The shortest stub is

$$\frac{l_2}{\lambda} = 0.1065 \qquad l_2 = 2.13 \text{ cm.}$$

Now consider the graphical solution. We enter the polar diagram at the normalized load admittance $Y_R/Y_0 = 0.3 + j0$ (point P_0 in Fig. 13) and observe that $t_0 = 0.30$ (on the αl circles) and $w_0 = 0$ (on the "wavelengths toward the generator" scale). As we move back toward the generator, the admittance point moves in a clockwise direction on the constant-αl circle. At P_1 we intersect the unity conductance circle and at the corresponding point on the transmission line we place the stub. At this point the admittance is $Y_1/Y_0 = 1 + j1.26$ and we have $w_1 = 0.17$. The distance l_1 from the load to the stub is determined by $l_1/\lambda = w_1 - w_0 = 0.17$.

The stub must provide a normalized susceptance of -1.26 mhos. Let us now determine the length of stub. At the short-circuited end of the stub we have $Y/Y_0 = \infty$. Entering the diagram at this point (Q_0 in Fig. 13), we observe that $w'_0 = 0.25$. Moving in the clockwise direction around the constant-αl circle we stop at the susceptance $Y_S/Y_0 = -j1.26$. The corresponding value of w' is $w' = 0.356$. Thus, the stub length is obtained from $l_2/\lambda = w' - w'_0 = 0.106$. The values of l_1/λ and l_2/λ obtained by the graphical method are in agreement with the analytical solutions.

10.11. Double-stub Impedance Matching.—In order to match variable load impedances using the single stub, it is necessary to vary the length of the stub as well as its position with respect to the load impedance. The stub position may be varied by inserting a telescoping "line stretcher" between the load and the stub. A more convenient impedance transforming system consists of two or more adjustable stubs spaced approximately a quarter wavelength apart as shown in Fig. 14.

Let us assume that $Y_g = Y_0$. To obtain maximum power transfer, it is then necessary to adjust the lengths of the stubs so that the admittance looking to the right at ab is equal to Y_0. Stub 1 serves to make the conductance part of the admittance at ab equal to Y_0, and stub 2 is then adjusted to cancel the susceptance portion of the admittance at ab.

Consider the graphical solution of the double-stub impedance transformer. Assume that the admittance Y_{cd} at points cd in Fig. 14 (with both stubs disconnected) corresponds to point P_0 in Fig. 15. Connecting stub 1 adds a pure susceptance, causing the admittance point to move on the constant-conductance circle to a new position P_1 which is determined by the length of the stub. If the stubs are a quarter wavelength apart,
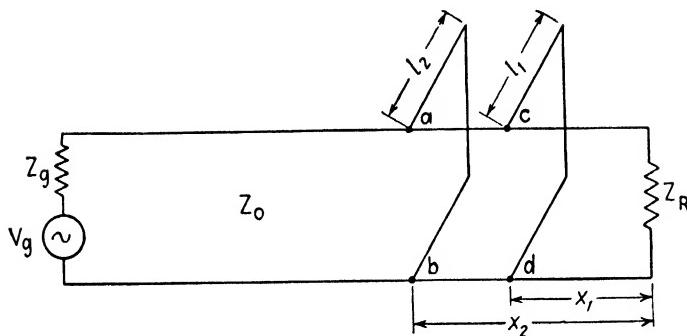
Fig. 14.—Double-stub impedance matching.
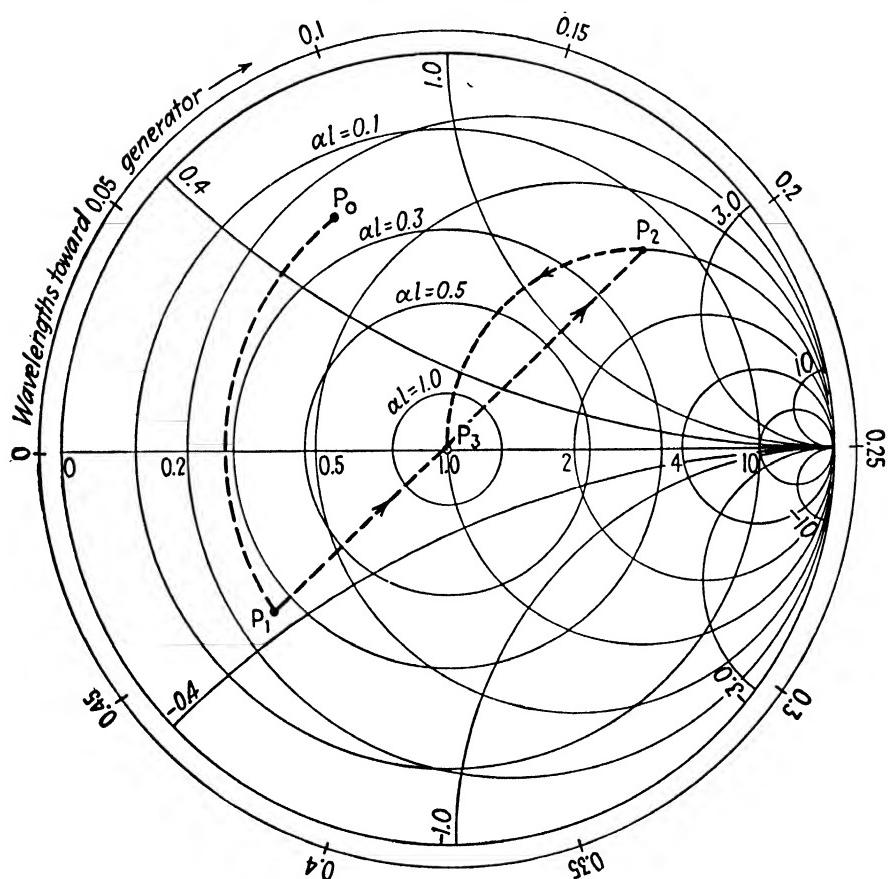
Fig. 15.—Graphical solution of double-stub impedance matching.

the admittance Y_{ab} (with stub 2 disconnected) is at point P_2 in Fig. 15, which is halfway around the diagram from P_1 on the same al circle.   In order to obtain the conditions necessary for matched admittances, the length of stub 1 should be such that point P_2 falls on the unity conductance circle.   With stub 2 connected and adjusted to cancel the susceptance at P_2, the admittance point moves from P_2 to P_3.   The latter point corresponds to $Y/Y_0 = 1 + j0$, which is the requirement for matched admittances.

The double-stub transformer will not match all possible load admittances. Thus, if the load admittance and position of the stub are such as to place
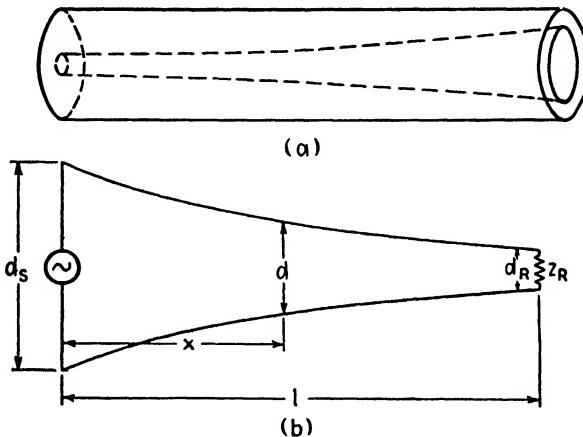
FIG. 16.—Coaxial and open-wire tapered lines.

P_0 on any conductance circle greater than unity (so that P_0 falls inside the unity conductance circle), it is then impossible to obtain matched admittances using two stubs which are spaced a quarter wavelength apart.   The range of admittances which can be matched by this method is increased somewhat by spacing the stubs a little less than a quarter wavelength apart.   Triple-stub impedance transformers are sometimes used where accurate impedance matching is required.

10.12. The Exponential Line.[1]—Tapered lines, which are several wavelengths long, such as those shown in Fig. 16, provide a gradual impedance transformation over the length of the line.   This type of impedance transformer is less frequency selective than the quarter-wavelength line or the stub transformers.   An exponential line is a tapered line in which the characteristic impedance varies exponentially along the line.

As a point of departure, let us write the differential equations of the transmission line similar to Eqs. (8.02-3 and 4), but with the distance x measured

[1] WHEELER, H. A., Transmission Lines with Exponential Taper, *Proc. I.R.E.*, vol. 27 pp. 65–71; January, 1939.

from the sending end.   For a lossless line these are

$$\frac{dV}{dx} = -j\omega L I \tag{1}$$

$$\frac{dI}{dx} = -j\omega C V \tag{2}$$

where the inductance L and capacitance C per unit length of line are now functions of the distance x.

We now assume that the tapered line is terminated in such a manner as to prevent reflections at the receiving end.   For this condition, there will be outgoing waves of voltage and current but no reflected waves. Assume that the outgoing wave of voltage has a variation with x given by

$$V = V_S e^{-(\delta + j\beta)x} \tag{3}$$

where V_S is the amplitude of the sending-end voltage and $e^{-\delta x}$ represents an exponential voltage transformation resulting from the line taper.   The constant δ will be referred to as the *voltage transformation constant.*

If the voltage is transformed by an amount $e^{-\delta x}$, we would expect that the current would be transformed by the inverse amount, or by an amount $e^{\delta x}$.   Hence the current is represented by

$$I = I_S e^{(\delta - j\beta)x} \tag{4}$$

where I_S is the sending-end current amplitude.

The voltage and current given by Eqs. (3) and (4) are in time phase at any point on the line and the power flow is

$$P = \tfrac{1}{2}|V||I| = \tfrac{1}{2}V_S I_S \tag{5}$$

The power is independent of distance x.

Let us now determine what conditions are required to obtain the assumed voltage and current distribution.   The impedance at any point along the line is the ratio of voltage to current.   Dividing Eqs. (3) and (4), we have

$$Z = \frac{V}{I} = Z_S e^{-2\delta x} \tag{6}$$

where $Z_S = V_S/I_S$.   Equation (6) shows that the impedance is transformed by the factor $e^{-2\delta x}$ which is the square of the voltage transformation.

Substitution of the voltage from Eq. (3) and the current from (4) into (1) and (2) gives

$$(\delta + j\beta)Z_S e^{-2\delta x} = j\omega L \tag{7}$$

$$-(\delta - j\beta)\frac{e^{+2\delta x}}{Z_S} = j\omega C \tag{8}$$

The product of these two equations gives

$$\delta^2 + \beta^2 = \omega^2 LC \tag{9}$$

and solving for β

$$\beta = \sqrt{\omega^2 LC - \delta^2} = \omega\sqrt{LC}\sqrt{1 - \frac{\delta^2}{\omega^2 LC}} \tag{10}$$

The phase constant β may be either real or imaginary, a condition similar to that encountered in filter networks. For large values of ω, the phase constant, given by Eq. (10), is real and the voltage and current waves propagate without attenuation (although they are transformed owing to the taper of the line). This corresponds to the pass band in conventional filter theory. For values of ω below a certain critical value, β is imaginary and the voltage and current, as given by Eqs. (3) and (4), are *both* attenuated with distance along the line. This corresponds to the attenuation band in filter theory. Hence the exponential line has properties similar to those of an impedance-transforming high-pass filter.

Cutoff occurs when β is zero, or from Eq. (10), when

$$\omega_c = \frac{\delta}{\sqrt{LC}} = \delta v_c \tag{11}$$

where ω_c is the cutoff angular frequency and $v_c = 3 \times 10^8$ meters per second is the velocity of light. Equation (11) shows that a high transformation constant δ results in a correspondingly high cutoff frequency.

The inductance and capacitance variation along the line are determined by the form of voltage and current distribution. Returning to Eq. (7) and writing this for the sending end ($x = 0$), we obtain $(\delta + j\beta)Z_S = j\omega L_S$ where L_S is the inductance per unit length at the sending end of the line. Dividing this into Eq. (7), gives

$$\frac{L}{L_S} = e^{-2\delta x} \tag{12}$$

A similar procedure applied to Eq. (8) yields

$$\frac{C}{C_S} = e^{2\delta x} \tag{13}$$

where C_S is the capacitance per unit length at the sending end.

If we define the characteristic impedance at any point on the line by the relationship $Z_0 = \sqrt{L/C}$, Eqs. (12) and (13) give

$$Z_0 = \sqrt{\frac{L}{C}} = \sqrt{\frac{L_S}{C_S}}\, e^{-2\delta x} \tag{14}$$

Finally, let us evaluate the terminal impedance Z_R which is necessary to prevent reflections. By dividing Eqs. (7) and (8), and substituting $Z_S = Ze^{2\delta x}$ from Eq. (6), we obtain

$$Z = \sqrt{\frac{L}{C}} \sqrt{\frac{-(\delta - j\beta)}{\delta + j\beta}} \qquad (15)$$

Rationalizing Eq. (15) and separating real and imaginary terms, we obtain

$$Z = \sqrt{\frac{L}{C}} \left(\frac{\beta}{\sqrt{\delta^2 + \beta^2}} + \frac{j\delta}{\sqrt{\delta^2 + \beta^2}} \right) \qquad (16)$$

Substitution of

$$\delta^2 + \beta^2 = \omega^2 LC = \left(\frac{\omega}{\omega_c}\right)^2 \delta^2 \qquad \text{and} \qquad \beta^2 = \omega^2 LC - \delta^2 = \delta^2 \left[\left(\frac{\omega}{\omega_c}\right)^2 - 1 \right]$$

from Eqs. (9) and (11) into Eq. (16) gives

$$Z = \sqrt{\frac{L}{C}} \left[\sqrt{1 - \left(\frac{\omega_c}{\omega}\right)^2} + j\frac{\omega_c}{\omega} \right] \qquad (17)$$

Equation (17) is an expression for the impedance (ratio of voltage to current) at any point on the exponential line. The impedance at either transmitting or receiving end of the line is obtained by inserting the values of L and C at the corresponding end of the line into Eq. (17). This gives the impedances which are required to terminate the line so as to prevent reflections.

If the impressed frequency is much greater than the cutoff frequency, then Eq. (17) reduces to $Z = \sqrt{L/C}$, i.e., the impedance at any point on the line is equal to the characteristic impedance at that point. For maximum power transfer, the generator and receiver impedances should then be approximately equal to the characteristic impedances at their respective ends of the line. In general, the condition $\omega \gg \omega_c$ is satisfied if the exponential line is several wavelengths long.

As the impressed frequency approaches cutoff the imaginary term increases. At cutoff the impedance is $Z = j\sqrt{L/C}$. The j signifies that the impedance is a pure reactance; hence there is no power flow.

Figure 17 is a plot of the voltage-transformation term $e^{-\delta x}$ and the impedance transformation term $e^{-2\delta x}$ against δx.

Example. An exponential line is desired to transform between resistive impedances $Z_g = 450$ ohms and $Z_R = 75$ ohms at a wavelength of $\lambda = 10$ cm. Find the length of coaxial line and the equation for the taper.

Inserting the values of Z_g and Z_R into Eq. (6), we obtain

$$\frac{Z_R}{Z_g} = e^{-2\delta l} = 0.167$$

Referring to Fig. 17, we obtain $\delta l = 0.9$. It is desirable to choose a length of line such that $\omega \gg \omega_c$. Inserting ω_c from Eq. (11), the inequality becomes $\omega \gg \delta v_c$ which may also be written $\delta \ll 2\pi/\lambda$. A value of δ equal to 10 per cent of $2\pi/\lambda$ would yield an impedance Z, which differs from Z_0 by less than 1 per cent. In order to obtain a reasonable length of tapered line, assume a value of $\delta = 0.06$. From the value of δl given above, we obtain $l = 15$ cm.
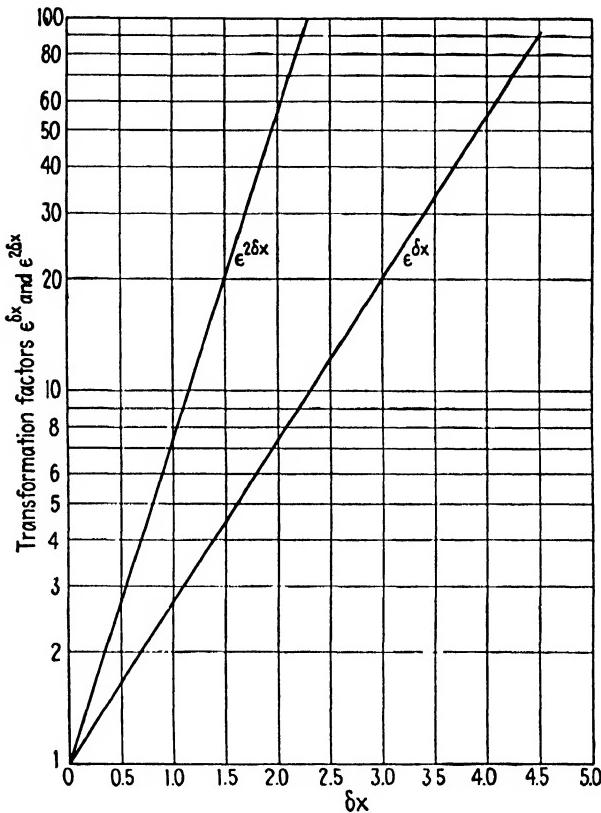
Fig. 17.—Plot of voltage and impedance transformation factors against δx.

The characteristic impedance of the line at the sending end should be approximately 450 ohms and that at the receiving end, approximately 75 ohms. To obtain the equation for the exponential taper, write Eq. (6) in the form $Z/Z_S = e^{-2\delta x}$. Inserting the characteristic impedance of a coaxial line from Table 1, Chapter 8, for Z, and $Z_S = 450$ ohms, we obtain

$$\frac{138}{450} \log_{10} \frac{b}{a} = e^{-2\delta x}$$

or

$$\log_{10} \frac{b}{a} = 3.26 e^{-0.12x}$$

where a and b are the radii of the inner and outer conductors, respectively.

10.13. Filter Networks Using Transmission-line Elements.[1]—Various types of filters may be constructed using transmission-line elements. In general, filter networks of this type are band-pass filters with multiple-pass bands due to the multiple-resonance properties of the component lines.

If the filter contains no lumped reactances, the resonant and antiresonant frequencies of the line elements are harmonically related and therefore the multiple-pass bands occur at harmonic intervals on the frequency scale. If lumped reactances are used, the multiple-pass bands, in general, will be inharmonically related.

(a)

(b)
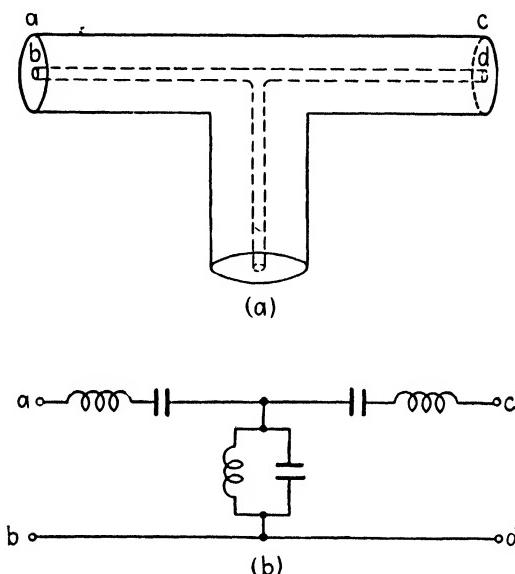
FIG. 18.—Band-pass filter and equivalent circuit

Consider the simple T filter shown in Fig. 18a. In the pass band, this may be represented by the equivalent circuit of Fig. 18b. The lengths l_1 and l_2 may be adjusted so that the series arms are resonant and the shunt arm is antiresonant at the mid-frequency in the first pass band. Let f_m represent this mid-frequency. At even harmonics of f_m the series arms are antiresonant and the shunt arm is resonant. This condition corresponds to maximum attenuation of the impressed signal. At odd harmonics of f_m the same conditions prevail as at the frequency f_m. Consequently, the pass bands are centered at odd harmonics of f_m, whereas the attenuation bands are centered at the even harmonics.

[1] MASON, W. P., and R. A. SYKES, The Use of Coaxial and Balanced Transmission Lines in Filters and Wide-Band Transformers for High Radio Frequencies, *Bell System Tech. J.*, vol. 16, pp. 275–302; July, 1937.

Figure 19a shows another type of band-pass filter using two resonators. These are joined by a transmission line which is effectively a quarter wavelength long at the mid-frequency of the first pass band. The resonators are tuned by adjustable plungers so as to be resonant at the mid-frequency of the pass band. In general, the multiple pass bands for this type of filter are not harmonically related since the capacitance between the end of the plunger and the bottom wall of the resonator is effectively a lumped reactance. Figure 19b shows the equivalent circuit for this type of filter. The width of the pass band is determined, in part, by the effective Q of the loaded resonators.
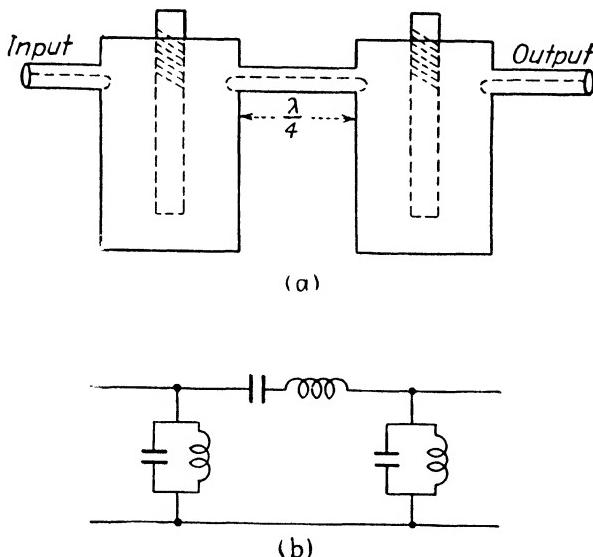
FIG. 19.—Band-pass filter and equivalent circuit.

Low-pass and high-pass filters may be constructed as shown in Fig. 20. The inductive reactances consist of short-circuited lines which are less than a quarter wavelength long.

A rigorous analysis of filters of the type shown in Figs. 18 to 20 is quite laborious. However, several useful relationships may be obtained by analogy with equivalent lumped-parameter networks. For example, the cutoff frequencies of low-pass and high-pass filters with lumped elements are $f_c = 1/\pi\sqrt{LC}$ and $f_c = 1/4\pi\sqrt{LC}$, respectively, where L and C are as shown in Figs. 20b and 20d.[1] The image impedance of the low-pass filter at zero frequency or the high-pass filter at infinite frequency is $Z_k = \sqrt{Z_1Z_2}$, where Z_1 is the total series impedance and Z_2 is the total

[1] EVERITT, W. L., "*Communication Engineering*," 2d ed., chap. 6, McGraw-Hill Book Company, Inc., New York, 1937.

shunt impedance of the lumped-parameter filter. The attenuation constant α for the lumped-parameter filter, in the attenuation band, is given by $\cosh \alpha = |\, 1 + Z_1/2Z_2\,|$. These relationships may be used as first approximations in the design of filters such as those described above. It is helpful to construct experimental filters with elements which can be adjusted to give the desired characteristics. The dimensions obtained from the experimental filter are then used in the design of the actual filter.
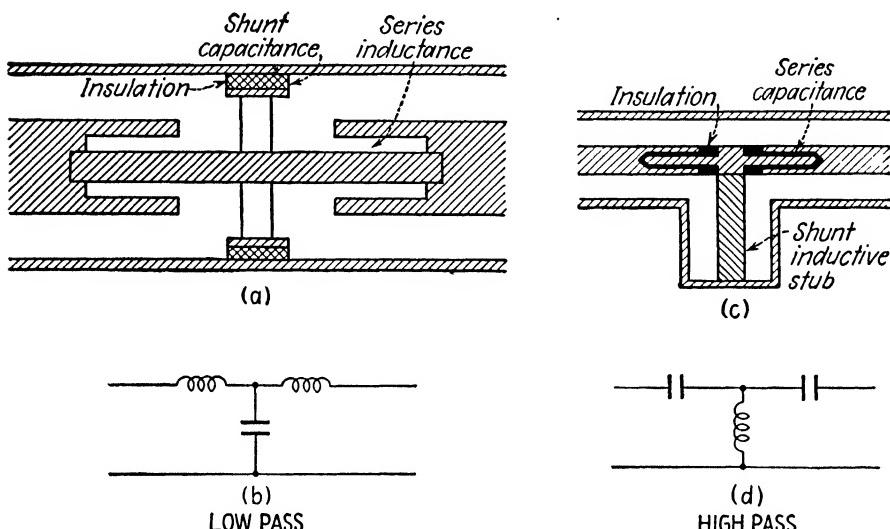
FIG. 20.—Low-pass and high-pass filters and their equivalent circuits.

In order to illustrate the methods of analysis of distributed-parameter filter networks, let us consider the band-pass filter shown in Fig. 18a. It is assumed that the transmission elements are lossless. The image impedance Z_k for the network as a whole is given by

$$Z_k = \sqrt{Z_{oc}Z_{sc}} \tag{1}$$

where Z_{oc} is the impedance looking into terminals ab with terminals cd open-circuited, and Z_{sc} is the impedance at ab with terminals cd short-circuited. The image impedance is the impedance which should be used to terminate the network in order to prevent reflection. It corresponds to the characteristic impedance in transmission-line theory. If the two series branches in Fig. 18a are identical, the network is symmetrical and the image impedance at ab is equal to that at cd. If the series branches are unequal, the network is unsymmetrical and the image impedances at ab and cd are unequal and the network is then an impedance-transforming filter.

The propagation constant $\Gamma = \alpha' + j\beta'$ for the entire filter network contains an attenuation constant α' and a phase constant β'. The propagation constant is related to the open-circuited and short-circuited impedances by

$$\cosh \Gamma = \sqrt{\frac{Z_{oc}}{Z_{oc} - Z_{sc}}} \tag{2}$$

We may replace Γ by $\alpha' + j\beta'$ and expand the hyperbolic cosine to obtain

$$\cosh (\alpha' + j\beta') = \cosh \alpha' \cos \beta' + j \sinh \alpha' \sin \beta' \tag{3}$$

For a dissipationless filter, α' is zero in the pass band and β' is either zero or π radians in the attenuation band. Equation (3) then reduces to
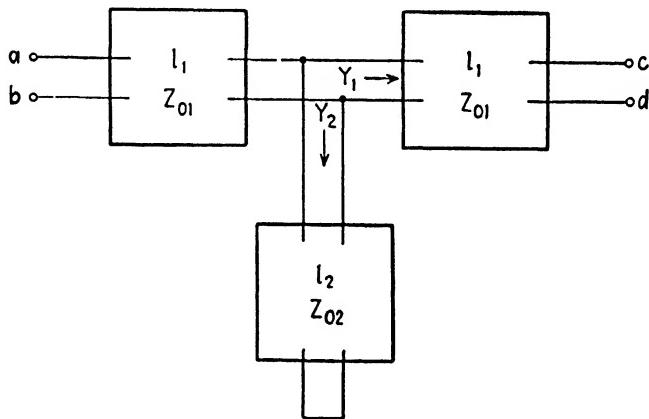
FIG. 21.—Block diagram representing the band-pass filter of Fig. 18a.

$\cosh \Gamma = \cos \beta'$ in the pass band and $\cosh \Gamma = \pm \cosh \alpha'$ in the attenuation band. In the band-pass filter, the pass band lies in the region defined by $-1 < \cosh \Gamma < +1$.

Let us now evaluate the image impedance and propagation constant for the filter shown in Fig. 18. A block diagram of this filter is shown in Fig. 21. Admittances will be used since we are dealing with parallel circuits. Consider first the admittance at terminals ab with cd short-circuited. The admittances Y_1 and Y_2 in Fig. 21 are

$$Y_1 = -jY_{01} \cot \beta l_1$$
$$Y_2 = -jY_{02} \cot \beta l_2 \tag{4}$$

The admittance $Y_1 + Y_2$ terminates the input branch of the filter. The input admittance of a lossless line terminated in an admittance Y_R may be obtained by writing Eq. (8.06-4) in terms of admittances, yielding

$$Y = Y_0 \left(\frac{Y_R + jY_0 \tan \beta l}{Y_0 + jY_R \tan \beta l} \right) \tag{5}$$

Substituting $Y_R = Y_1 + Y_2$, where Y_1 and Y_2 are given in Eq. (4), we obtain the admittance at ab with cd short-circuited,

$$Y_{sc} = jY_{01} \left[\frac{Y_{01}(\tan \beta l_1 - \cot \beta l_1) - Y_{02} \cot \beta l_2}{2Y_{01} + Y_{02} \tan \beta l_1 \cot \beta l_2} \right] \qquad (6)$$

Let us now find the input admittance at ab with cd open-circuited. With cd open-circuited, the admittances Y_1 and Y_2 are

$$Y_1 = jY_{01} \tan \beta l_1$$
$$Y_2 = -jY_{02} \cot \beta l_2 \qquad (7)$$

By substituting $Y_R = Y_1 + Y_2$ into Eq. (5), with the equations for Y_1 and Y_2 as given by Eq. (7), we obtain the open-circuited admittance

$$Y_{0c} = jY_{01} \left(\frac{2Y_{01} \tan \beta l_1 - Y_{02} \cot \beta l_2}{Y_{01} - Y_{01} \tan^2 \beta l_1 + Y_{02} \cot \beta l_2 \tan \beta l_1} \right) \qquad (8)$$

To obtain Z_k and $\cosh \Gamma$, substitute $Z_{sc} = 1/Y_{sc}$ and $Z_{0c} = 1/Y_{0c}$, where Y_{sc} and Y_{0c} are given by Eqs. (6) and (8), into Eqs. (1) and (2). Expressing the final result in terms of impedances, we obtain, after considerable manipulation,

$$Z_k = Z_{01} \sqrt{\frac{1 + (Z_{01} \tan \beta l_1 / 2Z_{02} \tan \beta l_2)}{1 - (Z_{01} \cot \beta l_1 / 2Z_{02} \tan \beta l_2)}} \qquad (9)$$

$$\cosh \Gamma = \cos 2\beta l_1 + \frac{Z_{01}}{2Z_{02}} \frac{\sin 2\beta l_1}{\tan \beta l_2} \qquad (10)$$

Now consider the special case in which $l_1 = l_2$, hence $\beta l_1 = \beta l_2$. Equation (10) may be written

$$\cosh \Gamma = \left(1 + \frac{Z_{01}}{2Z_{02}} \right) \cos 2\beta l + \frac{Z_{01}}{2Z_{02}} \qquad (11)$$

The mid-frequency of the pass bands occurs when the series branch is resonant and the shunt branch is antiresonant. In Sec. 8.06 it was shown that this occurs when $\beta l = n\pi/2$ where n is an odd integer. Inserting $\beta = \omega/v_c$ into this expression, the mid-frequencies are found to be

$$\frac{\omega_m l}{v_c} = \frac{n\pi}{2} \quad \text{or} \quad f_m = \frac{nv_c}{4l} \qquad n \text{ is odd}$$

where $v_c = 3 \times 10^8$ meters per second is the velocity of light. The condition that $\beta l = n\pi/2$, where n is odd, also corresponds to $\cosh \Gamma = -1$, as is evident by substitution of $\beta l = n\pi/2$ into Eq. (11).

Cutoff occurs when $\cosh \Gamma = 1$. Equation (11) then becomes

$$\cos 2\beta l = \frac{1 - (Z_{01}/2Z_{02})}{1 + (Z_{01}/2Z_{02})} \tag{12}$$

Comparing this with the identity $\cos 2\beta l = (1 - \tan^2 \beta l)/(1 + \tan^2 \beta l)$, we obtain

$$\tan \frac{\omega_c l}{v_c} = \pm \sqrt{\frac{Z_{01}}{2Z_{02}}}$$

$$\omega_c = \frac{v_c}{l} \tan^{-1} \pm \sqrt{\frac{Z_{01}}{2Z_{02}}} \tag{13}$$

The two cutoff frequencies for each band correspond to the positive and negative signs in Eq. (13). The width of the pass band decreases as the ratio $Z_{01}/2Z_{02}$ increases, approaching zero band width as $Z_{01}/2Z_{02} \to \infty$.
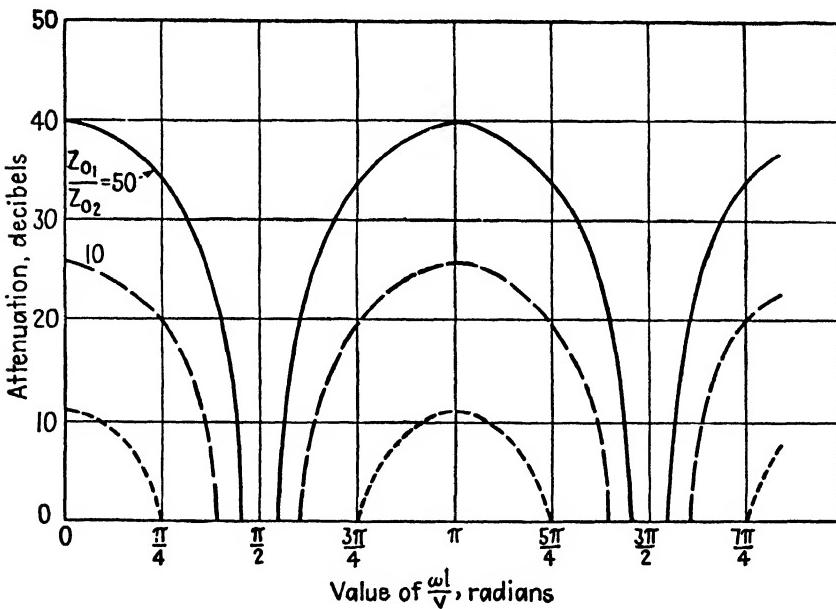
FIG. 22.—Attenuation characteristic of the filter shown in Fig. 18a.

From a practical point of view, the largest attainable ratio is of the order of $Z_{01}/2Z_{02} = 100$, corresponding to a minimum pass-band width of 20 per cent. In the attenuation band we have $\cosh \Gamma = \cos \alpha'$. A plot of α' as a function of βl for various values of $Z_{01}/2Z_{02}$ is shown in Fig. 22.

The image impedance at the mid-frequency is obtained by inserting $\beta l = n\pi/2$ (where n is odd) into Eq. (9), yielding

$$Z_k = \sqrt{1 + \frac{Z_{01}}{2Z_{02}}} \tag{14}$$

Now consider another special case in which it is assumed that $Z_{01} = 2Z_{02}$. Equation (10) then becomes

$$\cosh \Gamma = \frac{\sin \beta(2l_1 + l_2)}{\sin \beta l_2} \tag{15}$$

The mid-frequency for this filter occurs when $\cosh \Gamma = 0$, or when $\beta(2l_1 + l_2) = n\pi/2$, where n is even. Inserting $\beta = \omega/v_c$ into this expression, we obtain

$$f_m = \frac{nv_c}{4(2l_1 + l_2)} \qquad n \text{ is even} \tag{16}$$

Cutoff frequencies occur when $\cosh \Gamma = \pm 1$, corresponding to

$$f_{c1} = \frac{v_c}{4(l_1 + l_2)} \qquad f_{c2} = \frac{v_c}{4l_1} \tag{17}$$

The image impedance at the mid-frequency is

$$Z_k = Z_{01} \sqrt{-\tan\left(\frac{\pi/2}{1 + l_2/2l_1}\right) \tan\left[\frac{\pi}{2}\left(\frac{1 + l_2/l_1}{1 + l_2/2l_1}\right)\right]} \tag{18}$$

For narrow bands the image impedance is approximately $Z_k = (4l_1/\pi l_2)Z_{01}$.

The band width of this type of filter decreases as l_2 is made smaller. By making l_2 very small, it is possible to obtain a band-pass filter with a very narrow pass band.

PROBLEMS

1. A lossless line is terminated in a pure resistance which is not equal to the characteristic impedance of the line. Prove that the standing wave has its maximum and minimum values at the receiving end and at integral multiples of quarter-wavelength distances from the receiving end. Derive an expression for the standing-wave ratio for this case. Will the voltage be a maximum or a minimum at the receiving end?

2. A tuned circuit for an oscillator consists of an open-circuited line containing two different sizes of coaxial line as shown in Fig. 23. Derive an expression for the input impedance at terminals *ab* assuming that the lines are lossless. What are the conditions for antiresonance at *ab*? Show that the antiresonant frequency can be varied by varying the lengths l_1 and l_2, but keeping $l_1 + l_2$ constant. (*Note:* This is the principle used in tuning the lighthouse-tube oscillator shown in Fig. 6b, Chap. 5.)

3. A silver-plated coaxial line has dimensions $a = 0.5$ cm and $b = 2.0$ cm and is to be used at a frequency of 700 megacycles. Assume that the dielectric is lossless.

(*a*) Compute the attenuation constant and Q of the line.

(*b*) Evaluate the input impedance of a quarter-wavelength short-circuited section of line and a quarter-wavelength open-circuited section of line.

(*c*) Compute the data and plot a curve of the scalar value of input impedance as a function of frequence in the vicinity of the antiresonant impedance for the short-circuited line.

4. A line having a characteristic impedance of 75 ohms is terminated in an unknown impedance which is to be measured. The maximum and minimum voltages on the line are found to be 120 volts and 25 volts, respectively, with the maximum voltage point 30 cm from the terminal impedance. The frequency is 300 megacycles.

(*a*) Compute the value of the terminal impedance and check this value using the impedance diagram.

(*b*) Find the length and position of a single stub which will match the impedance to the line.
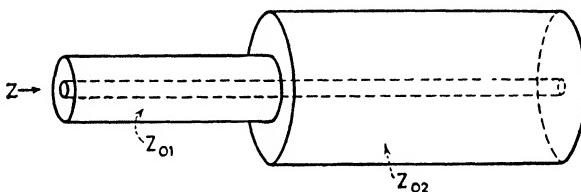
Fig. 23.

5. It is desired to construct an oscillator which will operate at 500 megacycles and deliver an appreciable amount of third harmonic in its output. A short-circuited coaxial line, having a characteristic impedance of 75 ohms, is used for the tuned circuit. This is shunted at its input end by the grid-plate capacitance of the tube which has a value of 1.5×10^{-12} farad. Determine the length of line and additional shunt capacitance which should be added at the input end of the line in order to make it simultaneously antiresonant at 500 and 1,500 megacycles.

6. The input circuit of a microwave receiver consists of a half-wavelength dipole antenna coupled to a coaxial line which is $2\frac{1}{2}$ wavelengths long. The line is terminated by a crystal detector. A short-circuited stub is placed near the receiving end in order to assure maximum power transfer to the detector and to increase the selectivity of the input circuit. The antenna impedance and the characteristic impedance of the line are both 72 ohms. The frequency selectivity of the circuit for crystal resistances of 150 ohms and 1,000 ohms are to be compared. The input circuit is tuned to a wavelength of 20 cm. The impedance diagram is to be used in the following calculations.

(*a*) Find the length of stub and its position for matched impedances at $\lambda = 20$ cm.

(*b*) Tabulate the values of complex impedance of the line at the antenna terminals (with the antenna disconnected) at wavelengths of $\lambda = 20 \pm 0.5k$ cm, where k is an integer from 1 to 10. Plot a curve of the scalar value of input impedance against wavelength.

(*c*) Assume that the antenna can be replaced by a generator having an internal impedance of 72 ohms. Using the impedances of part (*b*) and the power diagram of Fig. 9, plot curves of $P/P_{\max}$ against wavelength for the two cases considered.

(*d*) What conclusions can be drawn regarding the relative selectivity of the two systems?

7. Show that the power in a load impedance is given by

$$P = \frac{|V_{\max}||V_{\min}|}{Z_0}$$

where $|V_{\max}|$ and $|V_{\min}|$ are the maximum and minimum voltages on the transmission line supplying the load. This relationship is valid regardless of the magnitude or phase angle of the load impedance.
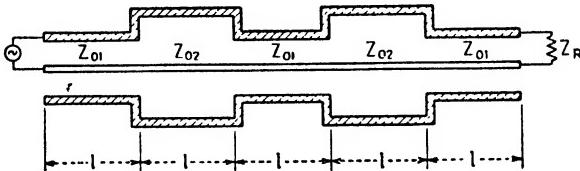
Fig. 24.

8. The system shown in Fig. 24 is a band-pass filter. The lengths l are all a half wavelength long in the middle of the pass band. Assume that $Z_R = 50$ ohms, $Z_{01} = 50$ ohms, and $Z_{02} = 200$ ohms. The wavelength at the middle of the pass band is 10 cm. Using the impedance diagram,

    (a) Find the input impedance of the system at wavelengths of $\lambda = 10 \pm 0.3k$ cm where k is an integer from 1 to 10.

    (b) Using the input impedances obtained in part (a) and the power diagram of Fig. 9, obtain the corresponding values of $P/P_{\max}$. Plot a curve of $P/P_{\max}$ against wavelength.

9. Design a symmetrical coaxial-line filter of the type shown in Fig. 18a to have a mid-frequency of 1,500 megacycles and a bandwidth of 5 megacycles. The image impedance is $Z_R = 150$ ohms at the mid-frequency. Plot a curve of α against βl for the filter.

CHAPTER 11

TRANSMITTING AND RECEIVING SYSTEMS

The fundamental processes involved in the transmission and reception of signals at microwave frequencies are essentially the same as those at ordinary radio frequencies. At the transmitter, the carrier may be generated by any one of the various types of microwave oscillators previously described, or it may be derived from a crystal oscillator followed by a chain of frequency multipliers. The carrier is modulated and the signal is then either impressed directly upon the transmitting antenna or it may be amplified by one or more successive stages of amplification before being radiated. Either amplitude, phase, or frequency modulation can be used. A new type of modulation, known as pulse-time modulation, also offers interesting possibilities at microwave frequencies.

Superheterodyne receivers are commonly employed in microwave systems. The input usually contains a frequency-selective circuit followed by a mixer. In the mixer, the incoming signal is heterodyned against a signal generated by a local oscillator to obtain the difference frequency. This difference frequency is amplified in one or more tuned stages of intermediate-frequency amplification after which it is detected and amplified in an audio- or video-frequency amplifier.

At microwave frequencies, difficulty is often encountered in maintaining a high degree of frequency stability of the carrier oscillator at the transmitter and of the local oscillator at the receiver. Various automatic frequency-control systems have been devised for the purpose of stabilizing these oscillators. Another difficulty arises owing to the fact that the band-widths of the microwave circuits and the tuned circuits in the intermediate-frequency amplifier are usually relatively large. Consequently these circuits admit a large amount of noise, resulting in a low signal-to-noise ratio. Certain types of microwave oscillators, particularly magnetrons, generate relatively high-noise voltages. These considerations favor the types of modulation in which the signal may be more readily separated from the noise, such as frequency modulation or pulse modulation.

A typical low-power microwave transmitting and receiving system, employing a frequency-modulated klystron oscillator and a superheterodyne receiver, is shown in the block diagram of Fig. 1. At the transmitter the klystron is frequency modulated by impressing the modulating voltage on its anode. Parabolic reflectors are used for directional transmission

Quarter wavelength
coaxial sleeve

Dipole
antenna

Parabolic
reflector

Frequency-
modulated
klystron

Audio-
frequency
amplifier

Input   (a) TRANSMITTER

Crystal
mixer

R-F Bypass
mica condenser

Plug for
removal
of crystal

Receiving
antenna

Local
oscillator
input

I.F.
amplifier
and
limiter

Discrim-
inator

Audio or
Video
amplifier

Output

Reflex klystron
local oscillator

(b) RECEIVER

Fig. 1.—Microwave communication system employing frequency modulation.

and reception. The local oscillator at the receiver consists of a reflex klystron oscillator, the output of which is combined with the incoming signal in the crystal mixer to obtain the intermediate frequency. The remaining portions of the receiver are similar to those of an ordinary frequency-modulation receiver.

11.01. Propagation Characteristics.—Long-distance radio communication is made possible by the reflection of radio waves from the ionosphere. The ionosphere consists of the upper atmosphere of the earth (from approximately 50 kilometers up), which is ionized principally by the passage of ultraviolet radiation from the sun through the rare atmosphere. The ionization has a tendency to become stratified, forming reflecting layers, the effective heights of which vary with the time of day, season of the year, sunspot activity, and other factors. Each layer has a more-or-less sharply defined cutoff frequency. At frequencies below the cutoff frequency the radio waves are reflected, whereas above the cutoff frequency they travel through the ionized layer without appreciable reflection. The low frequencies are reflected from the lower layers and the higher frequencies from the higher layers. The cutoff frequency for any one layer varies with the conditions of the ionized atmosphere.

Radio waves in the frequency band above 30 megacycles are seldom reflected from the ionosphere layers, and reflections above 150 megacycles are almost nonexistent. At still higher frequencies spurious reflections sometimes occur from the troposphere, which is that region extending from the earth's surface up to the ionosphere layer. These reflections are attributed to discontinuities in the dielectric constant at boundaries of air masses having different atmospheric characteristics. Occasionally the conditions are such as to form atmospheric "ducts" which have the properties of guiding the waves. These conditions make it possible to carry on microwave communication over distances exceeding the horizon distances. At the higher frequency end of the microwave band and in the infrared spectrum there are well-defined narrow absorption bands in which molecular absorption occurs. Communication over any appreciable distance in these absorption bands is virtually impossible.

Ground reflections or reflections from neighboring buildings may cause partial reinforcement or cancellation of the received signal, depending upon the relative phases of the direct and reflected waves at the receiver. In the transmission of television signals, these reflections may produce objectionable "ghost" images at the receiver. In general, it is desirable to locate the transmitting and receiving antennas as high as possible, with a clear line-of-sight path between the two antennas.

The power at the transmitter required to produce a given signal strength at the receiver can be greatly reduced by the use of directional·radiating systems. Consider a microwave transmitter with a total radiated power

of P_T watts.   If the power is radiated uniformly in all directions and there is no reflection or scattering of the waves, the power density (power per unit area perpendicular to the direction of propagation of the wave) at a distance r from the transmitter is $P_T/4\pi r^2$.

If a half-wavelength dipole antenna is used, the power is not uniformly radiated in all directions.   The power density at any point in a plane perpendicular to the antenna, passing through the center of the antenna, is then taken as

$$P = \frac{3P_T}{8\pi r^2} \tag{1}$$

The power absorbed by a half-wavelength dipole, used as a receiving antenna, corresponds to that in the equivalent area [1,2] $A_d = 3\lambda^2/8\pi$; hence the power at the receiver is

$$P_R = PA_d = \left(\frac{3\lambda}{8\pi r}\right)^2 P_T \tag{2}$$

If the transmitting antenna has a power gain of G_T as compared with a dipole antenna and the receiver is nondirectional, the receiver power is

$$P_R = \left(\frac{3\lambda}{8\pi r}\right)^2 P_T G_T \tag{3}$$

If the receiver also has a directional antenna with a power gain G_R, the receiver power is

$$P_R = \left(\frac{3\lambda}{8\pi r}\right)^2 P_T G_T G_R \tag{4}$$

Solving for the transmitter power, we obtain

$$P_T = \left(\frac{8\pi r}{3\lambda}\right)^2 \frac{P_R}{G_T G_R} \tag{5}$$

Hence, the transmitter power required for a given signal strength at the receiver varies inversely as the product of the gains of the transmitting and receiving antennas.

The power gain of a parabolic antenna may be approximated by the ratio of the effective absorption area of the parabola to that of the dipole. Let A_p be the area of the mouth of the parabola.   The effective absorption area can be taken as approximately $0.65A_p$, where the factor of 0.65 allows for the directional characteristics of the dipole antenna.   Dividing this

---

[1] SLATER, J. C., "Microwave Transmission," p. 244, McGraw-Hill Book Company, Inc., 1942.

[2] JENKS, F. A., Microwave Techniques, *Electronics*, vol. 18, pp. 120–127; October, 1945

effective area by A_d given above for the dipole antenna, the gain of the parabola is

$$G = 0.65 \times \frac{8\pi A_p}{3\lambda^2} \tag{6}$$

A typical parabola at a wavelength of 10 centimeters has an area of 6,000 square centimeters, resulting in a power gain of 327. If we compare two systems, one using dipole antennas at the transmitting and receiving ends and the other using dipole antennas with parabolic reflectors, we find that the power output of the first system would have to be approximately 100,000 times the power output of the second system in order to produce the same signal strength at the receiver. Hence, transmitters which are designed for general coverage require considerably greater power output than those designed for point-to-point communication.

TRANSMITTING SYSTEMS

11.02. Amplitude, Phase, and Frequency Modulation.[1]—Modulation is a process by which a high-frequency carrier wave is varied in accordance with the instantaneous value of the modulating voltage. In amplitude modulation, the envelope of the modulated wave has the same waveform as the modulating signal, as shown in Fig. 2a. In phase and frequency modulation, the amplitude of the modulated wave is constant, but its phase is shifted with respect to the unmodulated carrier, as shown in Fig. 2b. A phase shift is also accompanied by an instantaneous frequency shift; hence both the phase angle and the frequency of the modulated wave deviate with respect to those of the unmodulated carrier.

A modulated wave is approximately sinusoidal and may be represented by

$$v = V \cos(\omega_c t + \phi) \tag{1}$$

where V is the amplitude of the wave, ω_c is the angular frequency of the unmodulated carrier, and ϕ is a phase angle. In amplitude modulation, ω_c and ϕ are constant but V varies in accordance with the modulating signal. For sinusoidal modulation, we let $V = V_0(1 + m_a \sin \omega_m t)$, where m_a is the *amplitude-modulation factor* and ω_m is the angular frequency of the modulating voltage. Inserting this value of V into Eq. (1) and setting $\phi = 0$, we have

$$v = V_0(1 + m_a \sin \omega_m t) \cos \omega_c t \tag{2}$$

which may be expanded into

$$v = V_0 \cos \omega_c t + \frac{m_a V_0}{2} [\sin(\omega_c + \omega_m)t - \sin(\omega_c - \omega_m)t] \tag{3}$$

[1] EVERITT, W. L., Frequency Modulation, *Elec. Engineering*, vol. 59, pp. 613–625, November, 1940.

The amplitude-modulated wave therefore consists of a carrier wave and two side bands, the side bands having angular frequencies $\omega_c + \omega_m$ and $\omega_c - \omega_m$. If the modulating voltage is nonsinusoidal, it may be analyzed
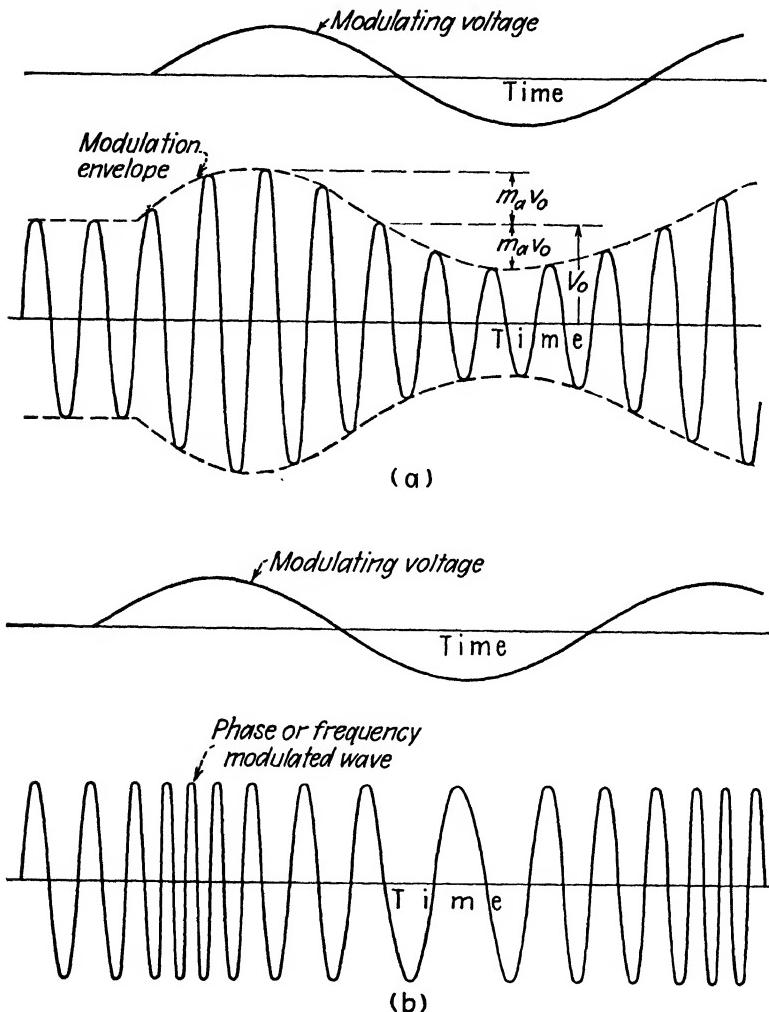
Fig. 2.—(a) Amplitude-modulated wave, and (b) phase- or frequency-modulated wave.

by Fourier series into sinusoidal components which are harmonically related. Each frequency component produces its corresponding side bands.

In phase modulation, the amplitude V in Eq. (1) is constant, but the instantaneous phase angle ϕ varies in direct proportion to the modulating voltage; hence we let

$$\phi = m_p \sin \omega_m t \tag{4}$$

where m_p is the *phase-modulation index*, representing the maximum phase-angle deviation from the unmodulated carrier in radians.   Let the instantaneous voltage be represented by

$$v = V_0 \sin (\omega t_c + \phi) \tag{5}$$

and insert ϕ from Eq. (4), to obtain

$$v = V_0 \sin (\omega_c t + m_p \sin \omega_m t)$$

$$= V_0[\sin \omega_c t \cos (m_p \sin \omega_m t) + \cos \omega_c t \sin (m_p \sin \omega_m t)] \tag{6}$$

The terms $\sin (m_p \sin \omega_m t)$ and $\cos (m_p \sin \omega_m t)$ may be represented by the following series containing Bessel functions

$$\sin (m_p \sin \omega_m t)$$

$$= 2[J_1(m_p) \sin \omega_m t + J_3(m_p) \sin 3\omega_m t + J_5(m_p) \sin 5\omega_m t + \cdots]$$

$$\cos (m_p \sin \omega_m t)$$

$$= J_0(m_p) + 2[J_2(m_p) \cos 2\omega_m t + J_4(m_p) \cos 4\omega_m t + \cdots]$$

Inserting these into Eq. (6), we have

$$v = V_0\{J_0(m_p) \sin \omega_c t + J_1(m_p)[\sin (\omega_c + \omega_m)t - \sin (\omega_c - \omega_m)t]$$

$$+ J_2(m_p)[\sin (\omega_c + 2\omega_m)t + \sin (\omega_c - 2\omega_m)t] \tag{7}$$

$$+ J_3(m_p)[\sin (\omega_c + 3\omega_m)t - \sin (\omega_c - 3\omega_m)t]$$

$$+ \cdots$$

Equation (7) shows that there is an infinite number of side bands having angular frequencies $\omega_c \pm n\omega_m$, where n takes integer values.[1]   However, if m_p (the maximum phase-angle deviation) is small, the side-band amplitudes diminish rapidly with increasing order of side-band frequency.   The relative side-band amplitudes for a phase-modulated signal with modulation frequencies of 2,000 and 10,000 cycles per second and a maximum phase deviation of $m_p = 4$ radians are shown in Figs. 3a and 3b.

A change in phase angle ϕ is also accompanied by a change in instantaneous frequency.   To show this, we return to Eq. (5) and let $\theta = \omega_c t + \phi$. The instantaneous angular frequency is

$$\omega = \frac{d\theta}{dt} = \omega_c + \frac{d\phi}{dt} \tag{8}$$

If we now insert ϕ from Eq. (4) into (8), we obtain

$$\omega = \omega_c + m_p\omega_m \cos \omega_m t \tag{9}$$

[1] Curves of $J_n(m_p)$ as a function of m_p are given in Fig. 2, Chap. 15.

The maximum deviation of the angular frequency is $m_p\omega_m$. Hence, in phase modulation, the maximum phase deviation m_p is independent of the modulating frequency, whereas the maximum angular frequency deviation $m_p\omega_m$ varies directly with the modulating frequency.

Frequency modulation is similar to phase modulation, except that the maximum frequency deviation is proportional to the amplitude of the modulating signal but is independent of its frequency. For sinusoidal modulation, the instantaneous angular frequency may be represented by

$$\omega = \omega_c + 2\pi \, m_f \cos \omega_m t \tag{10}$$

where m_f is the maximum frequency deviation in cycles per second. The phase angle ϕ may be obtained by equating (8) and (10), yielding

$$\phi = \int 2\pi \, m_f \cos \omega_m t \, dt = \frac{m_f}{f_m} \sin \omega_m t \tag{11}$$

Inserting this value of ϕ into Eq. (5), we obtain the equation of the frequency-modulated wave,

$$v = V_0 \sin \left(\omega_c t + \frac{m_f}{f_m} \sin \omega_m t \right) \tag{12}$$

Comparing Eqs. (6) and (12), we find that these are identical except for the substitution of m_f/f_m for m_p. Consequently, Eq. (7) may also be used as an expression for the frequency-modulated wave by the substitution of m_f/f_m for m_p.

The maximum phase-angle and frequency deviations for phase and frequency modulation may be summarized as follows:

	Maximum phase-angle deviation	Maximum frequency deviation
Phase modulation..........	m_p	$f_m m_p$
Frequency modulation......	$\dfrac{m_f}{f_m}$	m_f

In phase modulation, the maximum phase deviation is independent of the modulating frequency, whereas in frequency modulation, the maximum frequency deviation is independent of the modulating frequency. It should be noted that both m_p and m_f vary directly with the amplitude of the modulating signal.

Let us now compare the bandwidth of phase- and frequency-modulated signals. Assume that both the phase- and frequency-modulation systems have the same maximum frequency deviation at a modulating frequency of $f_m = 10,000$ cycles. Hence, when $f_m = 10,000$ cycles, we have $m_p = m_f/f_m$, and the side-band components, as expressed in Eq. (7), are identical for phase modulation and frequency modulation. These are represented by the bar graphs in Figs. 3a and 3c.

$f_m = 10,000$ cycles
$m_p = 4$ radians
$m_f = m_p f_m$
$\quad = 40,000$ cycles per sec.

(a) PHASE MODULATION (b)

$f_m = 2000$ cycles
$m_p = 4$ radians
$m_f = m_p f_m$
$\quad = 8000$ cycles per sec.

$f_m = 10,000$ cycles
$m_p = 4$ radians
$m_f = m_p f_m$
$\quad = 40,000$ cycles per sec.

(c) FREQUENCY MODULATION (d)

$f_m = 2000$ cycles
$m_p = m_f/f_m = 20$ radians
$m_f = 40,000$ cycles per sec.
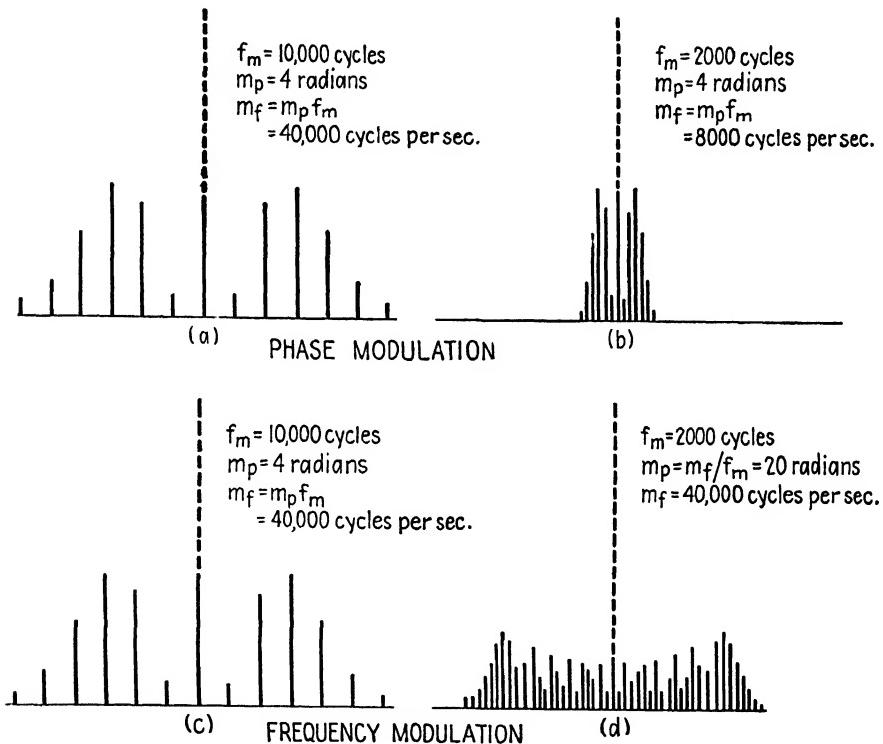
FIG. 3.—Comparison of sidebands in phase and frequency modulation. The maximum frequency deviation is assumed to be the same for both types of modulation at a modulating frequency of $f_m = 10,000$ cycles.

In the case of phase modulation, as the modulating frequency decreases, m_p remains constant and therefore the amplitudes of the side bands remain constant. However, as the modulating frequency decreases, the side bands move in closer to the carrier as shown in Fig. 3b. Consequently the effective bandwidth for phase modulation varies with the modulating frequency. We may look at this in another way. In phase modulation, the maximum frequency deviation is $f_m m_p$. Since this varies directly with f_m, it is clear that the maximum frequency deviation and hence the bandwidth both vary directly with the modulating frequency.

Now consider the bandwidth required for frequency modulation. In frequency modulation, we substitute m_f/f_m for m_p in Eq. (7). As the modulating frequency decreases, the side bands again move in closer to the carrier, but at the same time m_f/f_m increases; hence the number of significant side bands increases as shown in Fig. 3d. We find, therefore, that in frequency modulation, the effective bandwidth is substantially constant for all modulating frequencies. This conclusion can also be reached by noting that in frequency modulation, the maximum frequency deviation m_f is independent of the modulating frequency.

It is apparent from the foregoing discussion that the effective bandwidth varies directly with f_m for phase modulation but is independent of f_m for frequency modulation. The signal-to-noise ratio in either type of modulation is proportional to the bandwidth. Hence, for a given amount of frequency space, in general, frequency modulation results in a higher signal-to-noise ratio and is therefore preferred.

11.03. Methods of Producing Amplitude Modulation.—Amplitude modulation may be obtained by impressing the modulating voltage upon either the grid or the plate of a class C triode oscillator or amplifier. The modulation of an oscillator presents difficulties in that the carrier frequency shifts during the modulation cycle, thereby resulting in poor frequency stability. Also, oscillators will usually operate successfully over a limited range of grid or plate voltage; hence it is difficult to obtain a high percentage of modulation. The carrier frequency may be stabilized either by the use of automatic frequency control or by the use of compound modulation. In compound modulation the modulating voltage is impressed upon two electrodes simultaneously in such a manner as to tend to produce frequency changes in opposite directions, thereby maintaining constant carrier frequency.

In plate modulation, the class C stage is adjusted so that its radio-frequency output voltage varies directly with the plate-supply voltage over the range of modulation desired. When the modulating voltage is impressed upon the plate, the envelope of the modulated wave then has the same waveform as that of the modulating voltage. For 100 per cent modulation, the modulating power required is 50 per cent of the d-c plate power supplied to the class C stage during unmodulated conditions. The modulator must therefore be designed to deliver the necessary power output.

In grid modulation, the class C stage is adjusted for optimum operation at the least negative grid voltage which will be obtained during the modulation cycle. For undistorted modulation, the radio-frequency output voltage should vary linearly with grid voltage over the range of the modulating voltage.

Several systems of absorption modulation have been proposed as a means of producing amplitude modulation at microwave frequencies. In this type

of modulation, the modulating voltage varies an auxiliary load impedance which is coupled to the microwave oscillator or amplifier. The variations in load impedance result in corresponding variations in power absorption. The power delivered to the transmitting antenna is the difference between the power output of the oscillator or amplifier and that absorbed in the auxiliary load impedance. Hence, variations in power absorbed in the auxiliary load impedance result in inverse variations in power delivered to the antenna, thereby producing amplitude modulation. The auxiliary load can consist of a vacuum-tube circuit with the modulating voltage applied to the grid. The output of this modulator stage is coupled to the microwave system in such a way as to absorb microwave power in inverse proportion to the modulating voltage.[1]

11.04. Methods of Producing Phase and Frequency Modulation.—The frequency of most microwave oscillators can be varied over a range of several hundred kilocycles or more by varying the potential of an electrode in the oscillator. Frequency modulation can therefore be obtained by impressing the modulating voltage upon an electrode which is frequency sensitive to voltage. In order to produce undistorted modulation, it is necessary that the frequency vary linearly with the modulating voltage, while the power output remains constant. The linear frequency-voltage relationship is more important than constant power, since the limiter stage in the receiver serves to smooth out any amplitude variations which might occur. In most microwave oscillators, relatively large frequency deviations are produced by small changes in voltage, hence the modulating voltage can be relatively small.

The principal difficulty encountered in phase- or frequency-modulation systems is that of stabilizing the average frequency of the oscillator. Several automatic-frequency-control systems have been devised to stabilize the oscillator frequency without interfering with the modulation process. These are described in the following section.

An arrangement for frequency modulating a klystron oscillator is shown in Fig. 1. The modulating voltage is in series with the d-c resonator potential. Figure 14, Chap. 6 shows how the frequency and power output vary as a function of resonator potential. The power output is maintained reasonably constant over a relatively wide frequency range by tuning the resonators to slightly different frequencies.

The reflex klystrons can be frequency-modulated by inserting the modulating voltage in series with the reflector potential. The oscillator is then adjusted to obtain a linear frequency deviation as a function of the modulating voltage.

[1] RODER, Hans, Analysis of Load Impedance Modulation, *Proc. I.R.E.* vol. 27 pp. 386–395; June, 1939.

Another method of producing either phase or frequency modulation at microwave frequencies is to modulate a class C amplifier or oscillator at ordinary radio frequencies and use a chain of frequency multipliers to obtain the desired microwave frequency. It is then possible to use either a crystal-controlled oscillator, or an oscillator which has been stabilized by any one of the conventional automatic-frequency-control systems. Since a relatively high-frequency multiplication is required, the phase-angle or frequency deviation at the oscillator can be quite small. The final multiplier stage can be a klystron with a multiplication ratio as high as 15 in a single stage. This can be followed by a klystron power amplifier which delivers the required power output.
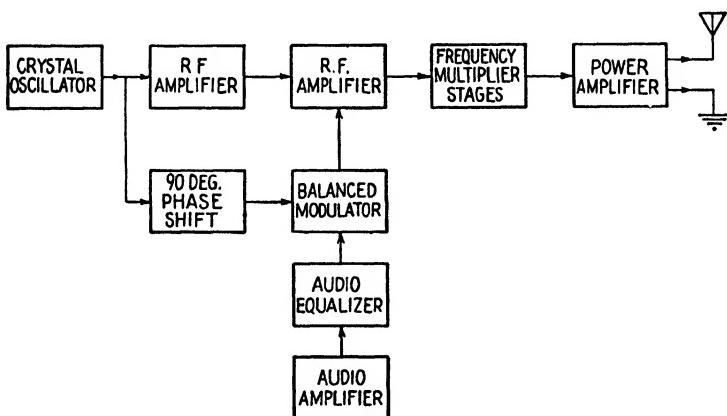
FIG. 4.—Armstrong method of producing frequency modulation.

The Armstrong method of producing frequency modulation at radio frequencies is represented in the block diagram of Fig. 4. This method is essentially a modified phase-modulation system. In Sec. 11.02, it was shown that in phase modulation the phase-angle deviation is independent of the modulating frequency, whereas in frequency modulation, it varies inversely with the modulating frequency. Hence, phase modulation can be readily converted into frequency modulation by merely inserting an R-C circuit into the audio-frequency channel, such that the output of this circuit varies inversely with the modulating frequency.

In the Armstrong system, the output of a crystal oscillator and the modulating voltage are both impressed upon the grids of a balanced modulator. The output of the balanced modulator contains the side bands corresponding to amplitude modulation but it does not contain the carrier. The carrier voltage, from the crystal oscillator, is shifted through a phase angle of 90 degrees and combined with the side bands from the balanced modulator.

In order to show how this process yields phase modulation, let us compare Eqs. (11.02-3 and 7). Assume that the maximum phase-angle deviation m_p is small in Eq. (7). We then have $J_0(m_p) \approx 1$, $J_1(m_p)$ is proportional to m_p, and all higher-order side bands are negligible and can therefore be discarded (see Fig. 2, Chap. 15). The principal distinction between Eq. (11.02-3) for amplitude modulation and (11.02-7) for phase modulation, then, is that the carrier-frequency term is a cosine term in the first equation and a sine term in the second. Amplitude modulation can therefore be converted into phase modulation by shifting the carrier through a phase angle of 90 degrees with respect to the side bands. This is exactly what is done in the Armstrong method. In this method it is necessary that m_p
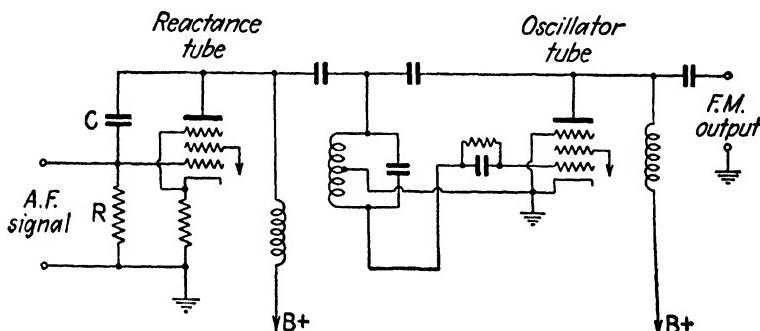
Fig. 5.—Reactance-tube modulator.

be less than one-half radian, in order to minimize distortion. By inserting the *R-C* circuit mentioned above into the audio-frequency channel, the modulation can be converted into frequency modulation.

As an example, assume that a crystal oscillator is operating at a frequency of 5 megacycles and is multiplied to a frequency of 3,000 megacycles, requiring a multiplication ratio of 600:1. If a maximum frequency deviation of 200 kilocycles is required at the 3,000-megacycle frequency, the corresponding maximum frequency deviation at 5 megacycles is $m_f = 200/600 = 0.333$ kilocycles. The corresponding phase-angle deviation for a modulating frequency of 1,000 cycles is $m_p = m_f/\omega_m = 0.0532$ radians.

The reactance-tube method of producing frequency modulation, shown in Fig. 5, consists of an oscillator with a reactance-tube circuit shunted across the tuned circuit of the oscillator. A reactance-tube circuit has the property of drawing a current which is approximately 90 degrees out of phase with the voltage across its terminals; hence it has the characteristics of a reactance. With the proper circuit constants, the effective reactance will be proportional to the transconductance of the tube. By

using a tube whose transconductance varies linearly with grid voltage and impressing the modulating voltage upon the grid, the effective reactance can be made to vary in direct proportion to the modulating voltage. This, in turn, produces an oscillator frequency deviation which is directly proportional to the modulating voltage, thereby producing frequency modulation. It is necessary to provide automatic frequency control for this system in order to stabilize the average carrier frequency of the reactance-tube modulator.

11.05. Automatic Frequency Control of Microwave Oscillators.—The purpose of automatic-frequency-control systems is to stabilize the frequency of an oscillator without interfering with the modulation process. Two principal methods are used: (1) comparison of the frequency with that
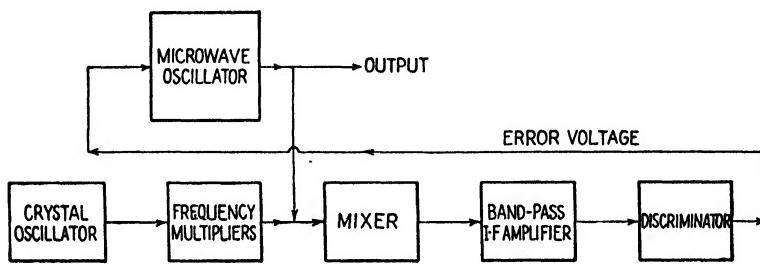
FIG. 6.—Block diagram of an automatic-frequency-control system.

of a standard oscillator and (2) comparison of the frequency with the resonant frequency of a standard tuned circuit. In either case, the frequency deviation is detected in a comparison circuit which feeds back an error voltage of the proper magnitude and polarity to correct the frequency of the oscillator.

The first method is illustrated by the block diagram of Fig. 6. In this system, the frequency of the microwave oscillator is compared with the output of a crystal-oscillator and frequency-multiplier chain. The difference frequency is amplified in a band-pass intermediate-frequency amplifier and detected in the frequency discriminator circuit. The discriminator circuit, shown in Fig. 12, has no output voltage as long as the impressed frequency (the intermediate frequency) is the same as the resonant frequency of the discriminator circuit. However, if the impressed frequency should deviate from this value (due to a frequency drift of the oscillator), the discriminator has an output voltage whose polarity and magnitude depend upon the direction and amount of frequency drift, respectively.

This error voltage from the discriminator circuit is impressed upon an electrode in the oscillator which is frequency-sensitive to voltage. By designing the discriminator circuit so as to have a long-time constant in comparison with the audio frequencies, it is possible to correct for slow

frequency drifts without interfering with the modulation.   The bandwidth of the feedback circuit and the loop gain of this circuit determine the degree of frequency stabilization.

Another system of automatic frequency control is shown in Fig. 7.   The microwave oscillator is frequency modulated by a "sensing" oscillator. The output of the microwave oscillator is impressed upon a high-Q resonator which is used as a reference frequency standard.   If the microwave frequency is the same as the resonant frequency of the standard resonator, the output of the resonator contains an amplitude-modulated wave which has
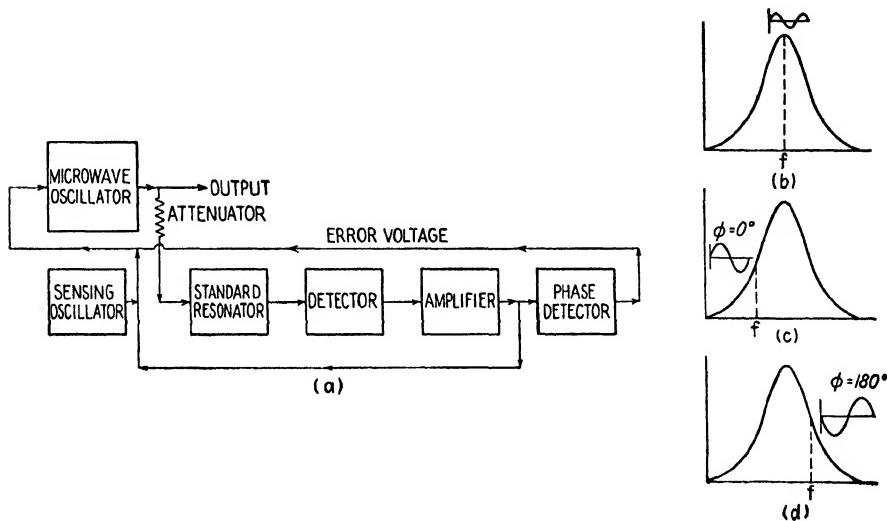
Fig. 7.—Block diagram of an automatic-frequency-control system.

twice the frequency of the sensing oscillator as shown in Fig. 7b.   If the frequency of the microwave oscillator is either above or below the resonant frequency, the output of the resonator is amplitude-modulated with a modulation frequency equal to that of the sensing oscillator, as shown in Fig. 7c and 7d.   There is a phase difference of 180 degrees in the amplitude-modulation envelopes, depending upon whether the impressed frequency is higher or lower than the resonant frequency of the standard resonator. The resonator output is detected and amplified, and the phase of the detected signal is compared with the voltage of the sensing oscillator in order to determine whether the microwave frequency is above or below the resonant frequency.   The phase-detector circuit feeds back an error voltage which corrects the oscillator frequency.   It is of course necessary to maintain accurate temperature control of the standard resonator.   The microwave oscillator can be amplitude-modulated in the customary way without interfering with the frequency-control system.

If the automatic-frequency-control system is to regulate the frequency of high-power oscillators, it may be necessary to feed the error voltage into a servomechanism which tunes the oscillator to maintain constant frequency.

RECEIVING SYSTEMS

11.06. Signal-to-noise Ratio in Receivers.[1]—The noise in microwave systems includes noises generated in the transmitter and receiver, and noises due to natural and artificial static. The most prominent of these are the noises generated in the receiver. These consist of: (1) noises due to the thermal agitation in resistances, including the radiation resistance of the antenna, (2) shot noises resulting from statistical fluctuations of electron emission, (3) secondary emission effects, and (4) miscellaneous noises. The noise voltages in the first three categories are more or less uniformly distributed over the entire frequency spectrum; hence the noise admitted by the tuned circuits in the receiver is proportional to the effective bandwidth of these circuits.

The mean square noise voltage due to thermal agitation in a resistance is $\mathbf{E}_n^2 = 4KTR_0 \, \Delta f$, where $K = 1.37 \times 10^{-23}$ joule per degree Kelvin is Boltzman's constant, R_0 is the resistance, T is the temperature in degrees Kelvin, and Δf is the bandwidth. If we consider $\mathbf{E}_n$ as being the noise voltage in the receiving antenna due to the random motion of particles in space, then R_0 is the effective resistance of the antenna.

Now consider the antenna as a generator, having an internal resistance R_0 and internal voltage $\mathbf{E}_n$ which is working into matched impedances, i.e., the load impedance is assumed to be R_0. The noise power delivered to the load is then $N_i = \mathbf{E}_n^2/4R_0$. Substitution of $\mathbf{E}_n$ from our previous relationship gives the noise power

$$N_i = KT \, \Delta f \qquad \text{watts} \qquad (1)$$

Although the equivalent temperature of space is not known, the value $T = 290$ degrees Kelvin (63 degrees Fahrenheit) has been suggested as a reasonable value. The noise power per cycle of bandwidth is then $N_i = KT = 4 \times 10^{-21}$ watt per cycle.

The signal power P_i at the input terminals of the receiver was given by Eq. (11.01-4). Dividing this by Eq. (1), we obtain the signal-to-noise power ratio at the receiver input terminals

$$\frac{P_i}{N_i} = \left(\frac{3\lambda}{8\pi r}\right)^2 \frac{P_T G_T G_R}{KT \, \Delta f} \qquad (2)$$

[1] FRIIS, H. T., Noise Figures of Radio Receivers, *Proc. I.R.E.*, vol. 32, pp. 419–422 July, 1944.

Additional noises are generated in the receiver. In order to account for these, a *noise figure* is defined as the ratio of the signal-to-noise ratio at the input terminals to the signal-to-noise ratio at the output terminals, assuming matched impedances at both the input and output terminals. Let N_0 be the noise power output, and P_0 be the signal power output for matched impedances. The noise figure F is then given by

$$F = \frac{P_i/N_i}{P_0/N_0} = \frac{P_i}{N_i} \times \frac{N_0}{P_0} \tag{3}$$

or the signal-to-noise ratio at the output terminals is

$$\frac{P_0}{N_0} = \frac{1}{F} \frac{P_i}{N_i} \tag{4}$$

Substituting P_i/N_i from Eq. (2) into (4), we obtain an expression for the signal-to-noise ratio at the output terminals,

$$\frac{P_0}{N_0} = \frac{1}{F} \left(\frac{3\lambda}{8\pi r}\right)^2 \frac{P_T G_T G_R}{KT \, \Delta f} \tag{5}$$

The noise figure is always greater than unity since P_0/N_0 is less than P_i/N_i. A noise figure of unity would correspond to an ideal receiver, *i.e.*, one which introduces no noise itself. This noise figure must be determined experimentally, since its value depends upon all of the noises generated in the receiver. In the design of communication systems it is customary to set a minimum signal-to-noise ratio which can be tolerated. The communication system must then be designed so as to produce a signal-to-noise ratio at the output terminals which exceeds this minimum value. Equations (4) and (5) show that the signal-to-noise ratio at the output terminals varies inversely as the noise figure. Hence, a reduction in receiver noises thereby decreasing the noise figure is just as beneficial as a proportionate increase in power output at the transmitter.

11.07. Frequency Converters.—The converter or mixer circuit is usually an integral part of the tuned input system, such as that shown in Fig. 1. In order to obtain a high degree of frequency selectivity of the input circuit, the crystal or diode impedance should differ appreciably from the antenna impedance. The stub matching system then produces an impedance match over a narrow range of frequencies and the input circuit is frequency selective.

In superheterodyne receivers there are two frequencies, equally spaced above and below the local oscillator frequency, which produce the same difference frequency. One of these is an undesired signal known as the *image frequency*. Another undesired signal which sometimes appears in the receiver is one having a carrier frequency equal to the intermediate fre-

quency of the receiver.   The rejection of these two unwanted signals must
occur in the tuned circuits preceding the converter.   The undesired signals
may be reduced by the use of tuned amplifier stages preceding the con-
verter.   At frequencies above approximately 1,000 megacycles, tuned
amplifiers have low gain and low signal-to-noise ratio; hence they are
not often used.

The frequency converter may be represented by the equivalent circuit
of Fig. 8a.   The incoming signal e_s and the local oscillator signal e_0 are
mixed in the crystal mixer which is assumed to have the ideal character-
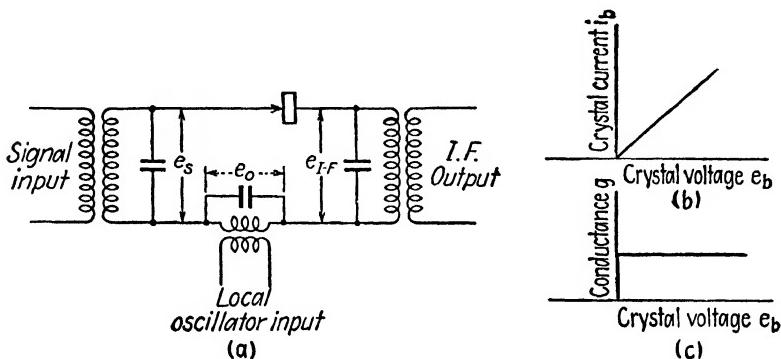
FIG. 8.—Crystal converter and characteristics.

istic shown in Fig. 8b.   The output circuit is tuned to the difference fre-
quency and is assumed to have negligible impedance at the frequency of the
incoming signal.

In the customary operation of the converter, the local oscillator voltage
e_0 is made considerably greater than the signal voltage in order to reduce
distortion.   For this condition, the conductance of the crystal at any instant
of time is dependent upon the local oscillator voltage.   If the local oscillator
voltage is $e_0 = E_0 \cos \omega_0 t$, the instantaneous conductance of the crystal can
be represented by the Fourier series

$$g = a_0 + a_1 \cos \omega_0 t + a_2 \cos 2\omega_0 t + \cdots a_n \cos n\omega_0 t$$

$$= a_0 + \sum_{n=1}^{\infty} a_n \cos n\omega_0 t \tag{1}$$

where the coefficients are determined by the shape of the conductance char-
acteristic of the mixer.

If we assume that the crystal resistance is high in comparison with the
impedance of the other circuit elements, the current will be determined by
the impedance of the crystal and may be represented by

$$i_b = ge_s \tag{2}$$

Now assume that the incoming signal is an unmodulated sine wave, given by $e_s = \mathbf{E}_s \sin \omega_s t$. Inserting this expression for e_s and the conductance from Eq. (1) into (2), the crystal current becomes

$$i_b = a_0 \mathbf{E}_s \sin \omega_s t + \mathbf{E}_s \sum_{n=1}^{\infty} a_n \cos n\omega_0 t \sin \omega_s t \qquad (3)$$

which may also be written

$$i_b = a_0 \mathbf{E}_s \sin \omega_s t + \frac{\mathbf{E}_s}{2} \sum_{n=1}^{\infty} a_n [\sin (\omega_s + n\omega_0)t + \sin (\omega_s - n\omega_0)t] \qquad (4)$$

The final term in Eq. (4) contains the intermediate-frequency term. The output circuit may be tuned to any one of the frequencies $\omega_s - n\omega_0$. Thus, conversion can take place with the fundamental frequency of the local oscillator ($n = 1$), or any harmonic of the fundamental frequency, corresponding to higher integer values of n. If the incoming signal includes modulation side bands, then the intermediate frequency would likewise include side-band terms having angular frequencies ($\omega_s - n\omega_0 + \omega_m$) and ($\omega_s - n\omega_0 - \omega_m$).

Equation (4) shows that the amplitude of the intermediate-frequency current of frequency ($\omega_s - n\omega_0$) is $I_{IF} = a_n \mathbf{E}_s / 2$. The *conversion conductance* for the nth harmonic of the local oscillator frequency is defined as this current divided by the signal voltage, or

$$g_{cn} = \frac{I_{IF}}{\mathbf{E}_s} = \frac{a_n}{2} \qquad (5)$$

The value of a_n may be obtained from the familiar expression for the coefficients of Fourier series,

$$a_n = \frac{1}{\pi} \int_0^{2\pi} g \cos n\omega_0 t \, d(\omega_0 t) \qquad (6)$$

where g is the conductance of the mixer. Since the value of g varies with the local oscillator voltage, g is a function of $\omega_0 t$.

The output voltage at the intermediate frequency is

$$\mathbf{E}_{IF} = I_{IF} Z_L \qquad (7)$$

where Z_L is the impedance of the output circuit at the intermediate frequency. The conversion gain is defined as the ratio of the output voltage at the intermediate frequency to the signal voltage. From Eqs. (5) and (7) we obtain the conversion gain

$$\frac{\mathbf{E}_{IF}}{\mathbf{E}_s} = \frac{a_n Z_L}{2} \qquad (8)$$

Thus far, the analysis holds for a mixer element having any form of $i_b - e_b$ characteristic.   Let us now evaluate the conversion gain for the ideal rectification characteristic of Fig. 8b.   The conductance $g = di_b/de_b$ is constant for positive values of voltage e_b and zero for negative values.

If a negative bias is placed on the tube, the variation of conductance with time is as shown in Fig. 9.   The conductance is zero until e_b becomes
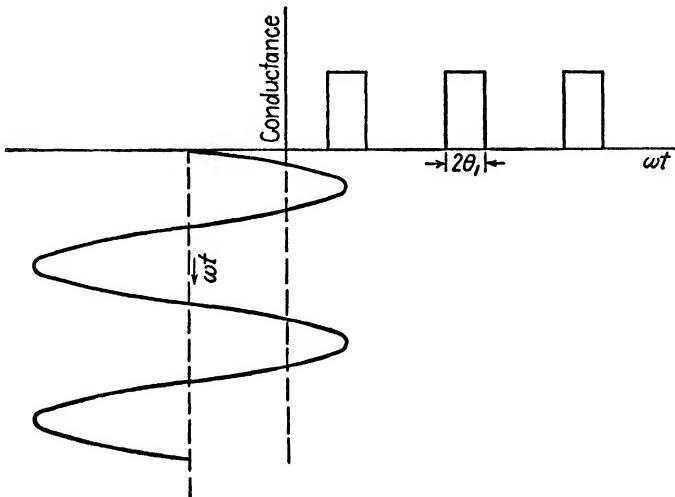
FIG. 9.—Conversion characteristic as a function of time for a crystal or diode mixer.

positive, then it suddenly rises to the value $g = di_b/de_b$ and is constant at this value until e_b turns negative.   If the conduction angle is from $-\theta_1$ to θ_1, the value of a_n in Eq. (6) becomes

$$a_n = \frac{g}{\pi} \int_{-\theta_1}^{+\theta_1} \cos n\omega_0 t \, d(\omega_0 t)$$

$$= \frac{2g}{\pi n} \sin n\theta_1 \tag{9}$$

Inserting Eq. (9) into (8), we obtain the conversion gain

$$\frac{\mathbf{E}_{IF}}{\mathbf{E}_s} = \frac{gZ_L \sin n\theta_1}{\pi n} \tag{10}$$

Maximum conversion gain occurs when $n = 1$ and $\theta_1 = \pi/2$, *i.e.*, when the intermediate frequency corresponds to $\omega_s - \omega_0$ and there is no bias voltage on the mixer.   The value of conversion gain for this condition is

$$\frac{\mathbf{E}_{IF}}{\mathbf{E}_s} = \frac{gZ_L}{\pi} \tag{11}$$

The conversion conductance g/π for this ideal condition is about one-third of the conductance of the crystal mixer.   The conversion conductance for higher harmonics is considerably smaller.

In order to obtain a high conversion gain, therefore, conversion should take place with the fundamental of the local oscillator frequency, the load impedance Z_L at the intermediate frequency should be as large as possible, and there should be no bias voltage on the mixer.   If the output circuit is broadly tuned, the value of Z_L decreases and the conversion gain is likewise reduced.

11.08. Intermediate-frequency Amplifiers.—In general, both the voltage gain and the signal-to-noise ratio of tuned amplifiers vary inversely with the bandwidth.   Therefore the objective in intermediate-frequency amplifier design is to obtain minimum bandwidth consistent with faithful amplification of the signal.   Certain types of modulation, such as that used in television, pulse modulation, and wide-band frequency modulation, require amplifiers having a relatively large bandwidth in order to reproduce the signal faithfully.

Wide-band amplification may be obtained by the use of: (1) doubletuned circuits which are tuned to the same frequency but overcoupled, (2) double-tuned circuits with primary and secondary tuned to slightly different frequencies, (3) the use of stagger tuning in which successive stages are tuned to slightly different frequencies, or (4) low-Q tuned circuits.

In the first method, two circuits are tuned to the same resonant frequency and are overcoupled.   This produces a voltage-gain curve which has two peaks, one on either side of the resonant frequency with a dip midway between the two peaks.   As the coefficient of coupling increases, the peaks separate farther apart and the dip at the center becomes more pronounced. In order to flatten out the over-all gain curve, the succeeding intermediate-frequency amplifier stage can be critically coupled so that its gain curve has a single peak midway between the peaks of the preceding stage.   This method requires high-Q circuits.

Stagger tuning of successive stages, *i.e.*, tuning the various stages to slightly different frequencies, produces approximately the same effect as overcoupling.

Low-Q tuned circuits offer an alternative method of obtaining wide-band amplifier characteristics.   The low Q is sometimes obtained by shunting the tuned circuit by a resistance.   For a given effective bandwidth, the voltage gain of the low-Q amplifier is less than that of the overcoupled amplifier.   Also, the overcoupled circuit has a higher attenuation outside the pass band.   However, the low-Q system is considerably easier to adjust in production.

11.09. Amplitude-modulation Detectors.—In simple types of microwave receivers, the signal may be fed directly into a crystal detector.   The

detected signal is then amplified by audio- or video-frequency amplifiers. In superheterodyne receivers, diode detectors are commonly used to detect the amplitude-modulated signals.

Amplitude-modulation detectors may be broadly classified into two categories: (1) square-law detectors and (2) linear detectors. The square-law detector operates on the principle that the detection characteristics of most detectors are not linear, but rather, are more nearly parabolic. The output voltage then contains frequencies corresponding to the sum and difference of the input frequencies. The difference frequency contains a voltage proportional 'to the original modulation voltage. The square-
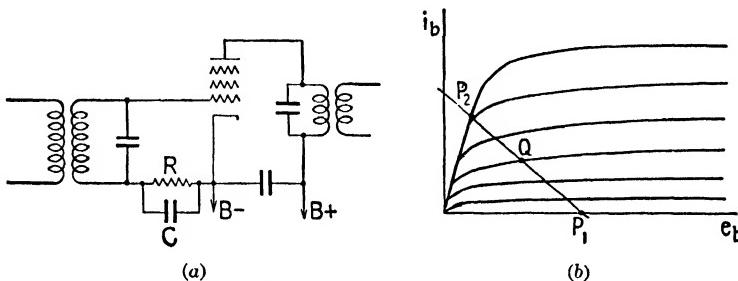
FIG. 10.—Intermediate-frequency amplifier used as a limiter.

law detector is essentially a small signal detector, since its characteristic approximates a parabolic (square-law) curve for only a limited range of operation.

In the linear detector, rectification of the wave takes place because of the rectifying properties of the detector. The envelope of the rectified current has the same waveform as the modulation voltage. An *R-C* circuit is used to take out the radio-frequency components, leaving a wave having the same waveform as the original modulation voltage. Linear detectors are essentially large-signal detectors.

11.10. Limiters and Discriminators in Frequency-modulation Receivers.—In frequency-modulation receivers, the incoming signal is reduced to a constant amplitude by a limiter stage before detection. This serves to reduce the noise voltage which appears largely as amplitude modulation in the signal.

An intermediate-frequency amplifier using a pentode with a low plate voltage may be used as a limiter. If a relatively large grid-driving voltage is used, the operating point varies between P_1 and P_2 on the load-line characteristic in Fig. 10b and the amplitude of the output voltage is substantially constant. The tuned circuit in the output serves to provide a sinusoidal output voltage, thereby eliminating waveform distortion. Further limiting action may be obtained by using a pentode with a remote cutoff characteristic and inserting a parallel resistance-capacitance combination in

the input circuit, as shown in Fig. 10a.   Rectified grid current flows through the R-C circuit, producing a negative bias voltage.   The bias voltage increases with an increase in signal strength, causing a reduction in the trans-
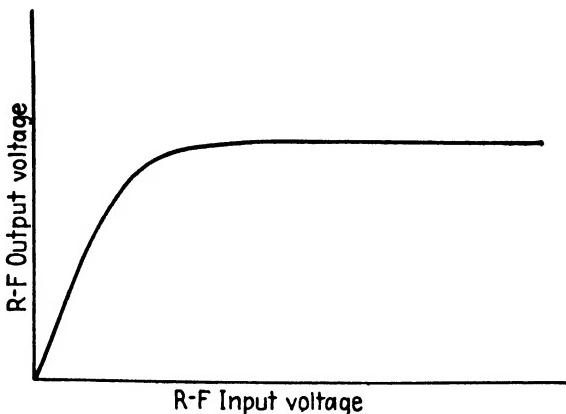
FIG. 11.— Characteristic of the limiter stage.

conductance of the tube.   The voltage gain of the stage then varies inversely with input voltage, thereby giving a constant output voltage. Figure 11 shows how the intermediate-frequency output voltage varies with input voltage for a typical limiter.
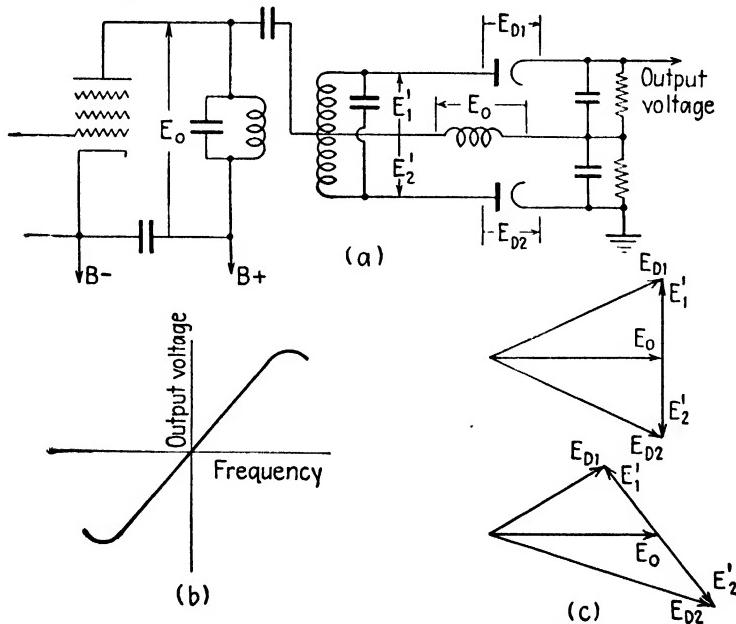
FIG. 12.—Discriminator circuit and characteristics.

A typical discriminator circuit for the detection of frequency-modulated signals is shown in Fig. 12. This consists of a tuned circuit which is connected to two diodes in such a manner that the output voltage is dependent upon the difference between the rectified currents passed by the two diodes.

In the circuit of Fig. 12, assume that the impedance of the R-C circuit is negligible at the intermediate frequency. The voltage across the two diodes is then $E_{D1} = E_1' + E_0$ and $E_{D2} = E_2' + E_0$, where E_1' and E_2' are each one-half of the voltage of the tuned circuit. If the impressed frequency is equal to the resonant frequency of the discriminator tuned circuit, the voltage relationships are those shown in the first diagram of Fig. 12c. The diodes then have equal impressed voltages and pass equal rectified currents. Since the currents flow in opposite directions in the load resistance, the output voltage is zero. If the impressed frequency differs from the resonant frequency of the discriminator, the diode voltages and diode currents are unequal, and there will be a net output voltage. The output voltage plotted against frequency is shown in Fig. 12c. An ideal discriminator would have a linear variation of output voltage with frequency.

CHAPTER 12

PULSED SYSTEMS—RADAR

In pulsed systems the transmitted wave consists of a succession of carrier pulses, each of very short duration. These are interspaced by relatively long time intervals during which there is no transmitted signal.

Prior to the war, pulsed systems were used, to a limited extent, to measure the height of the ionosphere layers and in a few experimental radar units. However, most of our present-day knowledge of pulsing systems and techniques can be attributed to the intensive wartime effort devoted to the
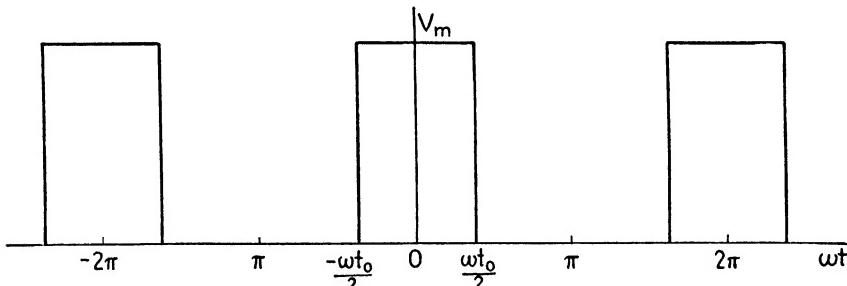
Fig. 1.—Rectangular pulse.

research and development of radar systems. Despite the fact that our present knowledge of pulsed systems is of comparatively recent origin, they have found widespread application throughout the field of communication and undoubtedly many interesting and useful applications will be discovered in the future.

12.01. Fourier Analysis of Rectangular Pulses.—Consider the rectangular pulse shown in Fig. 1. Let t_0 be the pulse duration, f be the repetition rate (number of pulses per second), and V_m be the peak value of the pulse. In the Fourier series analysis of the rectangular pulse, the fundamental frequency component is the pulse repetition rate f. Let $\omega = 2\pi f$ be the angular frequency. The Fourier series for the instantaneous voltage v of the wave of Fig. 1 contains only odd harmonic cosine terms, or

$$\text{v} = V_0 + \sum_{n=1}^{\infty} V_n \cos n\omega t \tag{1}$$

237

where n is any odd integer and V_n is the amplitude of the harmonic. The angular duration of the pulse is from $-\omega t_0/2$ to $\omega t_0/2$. The average value of voltage is

$$V_0 = \frac{V_m}{2\pi} \int_{-\omega t_0/2}^{\omega t_0/2} d(\omega t) = \frac{\omega t_0 V_m}{2\pi} \tag{2}$$

By the customary method of evaluating the Fourier series coefficients, we obtain the amplitude of the nth harmonic,

$$V_n = \frac{V_m}{\pi} \int_{-\omega t_0/2}^{\omega t_0/2} \cos n\omega t \, d(\omega t) = \frac{2V_m}{\pi n} \sin \frac{n\omega t_0}{2} \tag{3}$$
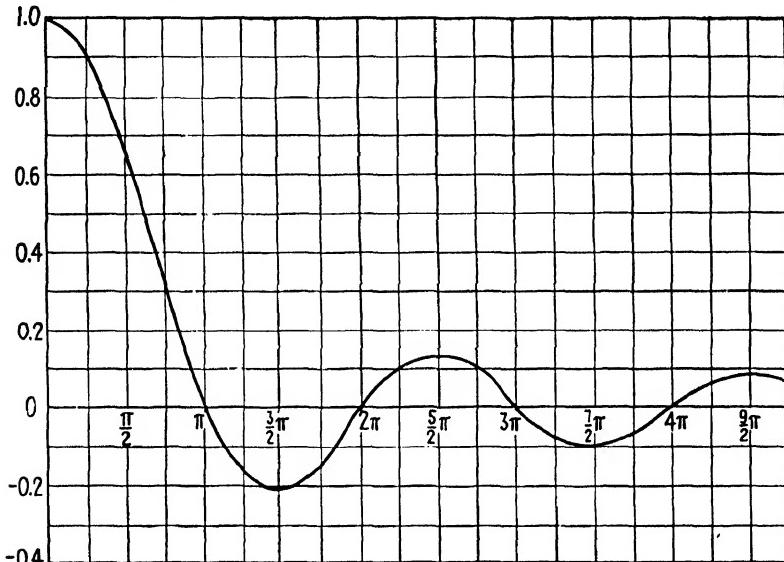
FIG. 2.—Plot of $\dfrac{\sin x}{x}$ as a function of x.

The ratio of the nth harmonic voltage to the average is found by dividing Eq. (3) by (2), yielding

$$\frac{V_n}{V_0} = 2 \frac{\sin (n\omega t_0/2)}{n\omega t_0/2} = 2 \frac{\sin x}{x} \tag{4}$$

where $x = n\omega t_0/2$. The curve of $\sin x/x$ as a function of x is plotted in Fig. 2. This curve may be used to determine the ratio V_n/V_0 for any harmonic component.

Now assume that the pulse duration is very small in comparison with the time interval between pulses. For the fundamental frequency component, we have $n = 1$; also $x = \omega t_0/2$ is very small. The corresponding value of $\sin x/x$, from Fig. 2, is approximately unity. Hence, Eq. (4)

shows that the fundamental frequency component has an amplitude equal to twice the average value. In order to observe how rapidly the amplitudes of the harmonics decrease with the order of the harmonic, let us find the harmonic which has an amplitude equal to 0.707 of the fundamental amplitude. Referring to Fig. 2 we observe that $\sin x/x$ has a value of 0.707 when $x = n\omega t_0/2 = 1.46$ radians. Solving for n we obtain

$$n = \frac{2.92}{\omega t_0} \qquad (5)$$

The effective bandwidth nf will be defined as the bandwidth which includes all harmonic components having magnitudes greater than 0.707 of the fundamental. Multiplying both sides of Eq. (5) by f, we obtain the bandwidth as

$$nf = \frac{0.465}{t_0} \qquad \text{cycles per second} \qquad (6)$$

Hence, the effective bandwidth for pulses of very short duration varies inversely with the pulse duration but is independent of the repetition rate.

As an example, assume that a rectangular pulse has a duration of 1 microsecond and a repetition rate of 1,000 pulses per second. We then have $t_0 = 10^{-6}$ second and $f = 1,000$ cycles per second. The bandwidth is then $nf = 0.465/10^{-6} = 465,000$ cycles per second. If the rectangular pulse is used to modulate a carrier wave, the various harmonic components produce amplitude-modulation side bands with frequencies above and below the carrier frequency; hence the total effective bandwidth is twice the value given above, or 930,000 cycles.

The foregoing example shows that an extremely large bandwidth is required to transmit and receive pulses of very short duration. If the transmitting and receiving circuits have insufficient bandwidth to pass the significant harmonic components, then the pulse will be rounded off, thereby altering the waveform. The effective bandwidth may be reduced by using a rounded or triangular-shaped pulse. However, in many applications a steep wavefront is required in order to trigger circuits at a definite instant of time. In such cases the larger bandwidth must be tolerated.

12.02. Radar Principles.—Radar systems are used to locate a target and to determine accurately its position in space with respect to the radar unit. Pulsed waves are used as a means of determining the distance from the radar unit to the target, this distance being known as the range. The radar transmitter sends out pulsed waves which are partially reflected from the target. A small portion of the reflected energy returns to the receiver. Since the waves travel through space with a velocity equal to the velocity of light, the time interval between the transmission of a pulse and the recep-

tion of the reflected pulse is directly proportional to the range. Each microsecond of time delay corresponds to an increase in range of 164 yards.

In most radar systems the transmitter and receiver are assembled as a unit, with a single antenna serving both the transmitter and receiver. Highly directional antenna systems are usually used; hence the azimuth and elevation of the target may be determined by the position of the antenna when the reflected signal is a maximum. The range, azimuth, and elevation determine the position of the target with respect to the radar unit. In some applications, such as ship detection, it is not necessary to know the elevation.

Several standard scanning systems have been devised to enable the radar unit to search over a relatively large area. One of these is a circular scanning system in which the antenna revolves slowly and the radiated beam sweeps out a circular path. This supplies range and azimuth information only. In another system, the circular motion is combined with a slow vertical tilting motion so that the beam follows a helical path. A typical system might have a circular sweep at the rate of 6 revolutions per minute, with 4 degrees of vertical tilt for each revolution. This scanning system makes it possible to determine range, azimuth, and elevation.

Still another scanning system uses a conical sweep. This is obtained by using a dipole antenna and a parabolic reflector. The dipole antenna is a short distance from the focal point of the parabola and is rotated in a small circle which has the focal point at its center. This causes the radiated beam to sweep out a conical path with a cone angle which is dependent upon the distance between the dipole antenna and the focal point of the parabola. In some radar units, the antenna is equipped with motor drive and servomechanism control which permits the antenna to automatically track the target.

At the receiver, the range, azimuth, and elevation information usually appear on one or more cathode-ray oscilloscopes. Several types of sweep circuits have been devised in order to translate the information into a form which can be quickly interpreted. The simplest arrangement is the type *A* scope which has a saw-tooth wave applied to the horizontal deflection plates to provide a linear horizontal time axis. The pattern on the oscilloscope is as shown in Fig. 3a, the distance between the transmitted and reflected pulses being directly proportional to the range.

The type *J* oscilloscope, shown in Fig. 3b, has the outgoing and reflected pulses superimposed upon a circular trace. The angle between the two pulses is proportional to the range.

The PPI (plan-position indicator) oscilloscope, of Fig. 3c, has a trace which starts from the center of the oscilloscope, each time that a pulse is transmitted, and moves radially outward. A circular motion is superimposed upon the radial sweep so that the complete path resembles a wagon

wheel with a very large number of spokes.   The output of the radar receiver is impressed upon the grid of the oscilloscope, thereby controlling the beam current in the oscilloscope beam.   A bright spot on the screen denotes a reflection.   When used with the circular-scan antenna, the radial distance from the center of the oscilloscope screen to a bright spot (denoting a reflection) is proportional to the range of the target and the angular position is proportional to the azimuth.

The type B oscilloscope is similar to the PPI oscilloscope except that rectangular coordinates are used instead of polar coordinates.   The cathode-ray beam starts at a base line each time that a pulse is trans-
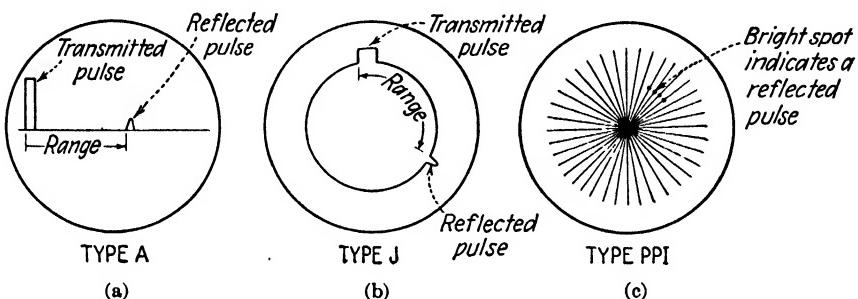
FIG. 3.—Oscilloscope traces for radar systems.

mitted and moves vertically during the interval between pulses.   The vertical distance is proportional to range and the horizontal distance is proportional to azimuth.

12.03. Specifications of Radar Systems.[1]—The specifications of a radar system, *i.e.*, the power output, pulse repetition rate, pulse duration, type of antenna, receiver sensitivity, etc., must necessarily be governed by the use for which the particular system is intended.   The following are some general principles which affect the choice of these specifications.

First, consider how the maximum range varies with transmitter power. In Sec. 11.01, it was shown that the power density of a wave radiated from a transmitting antenna varies inversely as the square of the distance.   A small portion of this power is intercepted by the target and is reradiated into space.   The power density in the reradiated wave likewise varies as the inverse square of the distance.   Consequently the power received by the radar receiver varies as the inverse fourth power of the range.   Because of this inverse fourth-power relationship, a relatively large increase in transmitter power results in a disproportionately small increase in maximum range.   For example, in order to double the maximum range of a radar system, it is necessary to increase the transmitter power by a factor of 2^4 or 16 times, assuming that all other factors are held constant.   In general,

[1] SCHNEIDER, E. G., Radar, *Proc. I.R.E.*, vol. 34, pp. 528–580; August, 1946

the maximum range is determined by the transmitter power, directivity of the antenna system, receiver sensitivity, and signal-to-noise ratio at the receiver.

Now consider the factors affecting the choice of pulse duration and pulse repetition rate. Since, in pulsed systems, the transmitter is usually in operation for a very small fraction of the total time, it is possible to obtain very high values of peak power without exceeding the safe average plate dissipation of the transmitter tubes. If we assume that the efficiency of a tube is the same for pulsed operation as for continuous-wave operation, the ratio of pulsed power output (peak value) to continuous-wave power output for the same average plate dissipation is

$$\frac{P_{\text{pulsed}}}{P_{cw}} = \frac{T}{t_0} \tag{1}$$

where t_0 is the duration of the pulse and T is the time interval between the beginning of one pulse and the beginning of the next succeeding pulse. For example, consider a system in which the pulse is 1 microsecond long and the repetition rate is 1,000 pulses per second ($T = 10^{-3}$ second). We then have $P_{\text{pulsed}}/P_{cw} = 10^{-3}/10^{-6} = 10^3$, or the safe peak power output available from the pulsed system is approximately 1,000 times the safe power output with continuous operation. The same average power is obtained in both cases.

From a casual observation, it would appear that the maximum range of a radar unit can be increased by decreasing the pulse duration and proportionately increasing the transmitter power output. However, if we investigate the conditions at the receiver, our conclusions are somewhat different. At the receiver, the bandwidth required to amplify the pulse without distortion varies inversely with the duration of the pulse. Hence, a shorter pulse requires a larger bandwidth which also results in an increase in noise voltage. The minimum discernible signal at the receiver is one whose voltage is approximately equal to the noise voltage; in other words, one which corresponds to a signal-to-noise ratio of approximately unity.

Thus, the shorter pulse makes higher peak power possible, but it also necessitates a larger bandwidth at the receiver, which increases the receiver noise. Consequently, the signal-to-noise ratio at the receiver is approximately the same for a short pulse with high peak power as for a long pulse with low peak power, assuming the same average power in both cases. We therefore conclude that, for a given average power output, the maximum useful range of a radar unit is independent of the pulse width. However, if the radar system is used for short ranges, a short pulse is required in order that the transmitter be off and the receiver fully recovered when the reflected signal arrives at the receiver.

If the radar system is intended for long-distance detection, a low pulse repetition rate is required, since the time interval between pulses must be sufficient to allow the reflected pulse to return before the next succeeding pulse is sent out. On the other hand, in order to have a good signal at the receiver, at least five pulses must hit the target each time that it is scanned. This may be accomplished by using a slow scanning speed or a high pulse-repetition rate. Pulse-repetition rates of typical radar systems range from 200 to 4,098 pulses per second and the pulse duration varies from 0.25 to 30 microseconds.
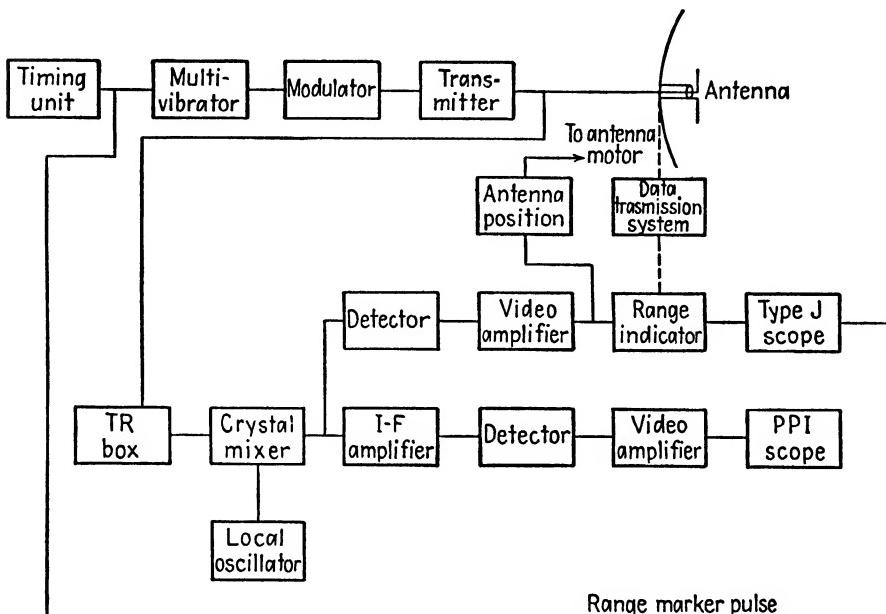
FIG. 4.—Block diagram of SCR 584 radar system.

12.04. Typical Radar System.—A simplified block diagram of the SCR-584 radar unit is shown in Fig. 4. This system was designed to direct the fire of antiaircraft batteries. It uses a pulsed magnetron transmitting tube operating at a wavelength of 10 to 11 centimeters and delivering a peak power output of 300 kilowatts. The pulse repetition rate is 1,707 pulses per second and the duration is 0.8 microseconds. When searching for a target the antenna describes a helical path at the rate of 6 revolutions per minute. The unit is equipped to automatically track a target, in which case conical scanning is used. The angular accuracy is ±0.06 degree and the range accuracy is ±25 yards.

The timing unit, shown at the extreme left, contains a quartz-crystal oscillator, frequency-dividing circuits, and pulse-shaping circuits. This

timing unit initiates the transmitter pulse and also serves to synchronize the oscilloscope as well as to provide marker pulses for accurate ranging. In the transmitter circuit, the timing circuit triggers a multivibrator at the beginning of each pulse. The multivibrator circuit contains an artificial transmission line, known as a *delay line*, which controls the duration of the pulse.

The modulator circuit, shown in Fig. 5, contains a condenser which is charged to a high voltage (22 kilovolts) during the period between pulses. This condenser is discharged by a modulator tube. The grid of the modu-- lator tube is normally biased beyond cutoff but is driven into the positive
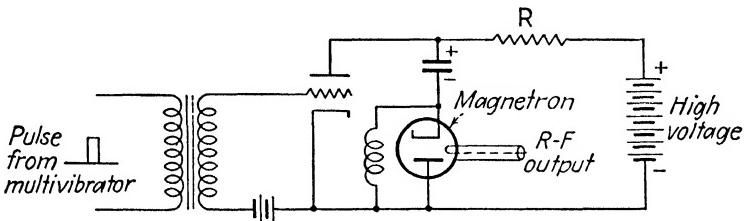
FIG. 5.—Simplified diagram of a pulse-modulator circuit.

region by the pulse from the multivibrator. This causes the condenser to discharge through the magnetron and the modulator tube. The high voltage impressed upon the magnetron causes it to oscillate at its natural frequency. The microwave power obtained from the magnetron is impressed upon the antenna.

The T-R (transmit-receive) box, shown in Figs. 4 and 6, serves to prevent the transmitted pulse from entering the receiver, without interfering with the reception of the reflected pulse. It contains a T-R tube, a tunable resonant cavity, and provision for coupling the input and output circuits to the cavity. The T-R tube has two conical metallic electrodes with the apexes separated a short distance apart. The tube is enclosed in a glass envelope and contains a slight amount of water vapor. The transmitter pulse causes a spark discharge in the T-R tube which detunes the resonant cavity. This introduces a very high attenuation between the transmitter and receiver circuits. At the conclusion of the transmitted pulse, the T-R tube deionizes and the passage is clear for the reflected signal to pass through the T-R box to the receiver.

The receiver contains a crystal mixer, a local oscillator consisting of a reflex klystron, seven stages of intermediate-frequency amplification, a detector, and video amplifier stages. One channel of the receiver output is impressed upon two type *J* oscilloscopes which are used in a range-indicating system. One oscilloscope has a range of 32,000 yards for coarse adjustment and the other 2,000 yards for fine adjustment. The operator

adjusts a range handwheel so as to keep the reflected pulse on a hairline on the oscilloscope, and this handwheel circuit feeds the range information into the data transmission system.

Another receiver channel is used to automatically position the antenna. The antenna uses a conical scanning system when tracking a target, and the antenna is positioned so that the reflected pulses are equal throughout the entire cycle of conical scan.   If the antenna position is slightly in error,
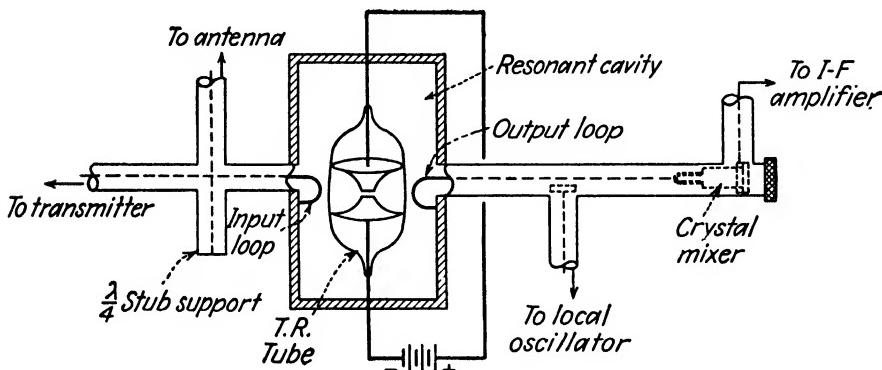
FIG. 6.—T-R box and crystal mixer.

the reflected signals are unequal during different portions of the conical scan.   An error signal is then fed into the servomechanism which operates to correct the position of the antenna.

The third receiver channel goes to the PPI oscilloscope which also receives trigger pulses from the timing unit.

12.05. Pulse-time Modulation.[1,2]—Several communication systems have been devised which use pulsed waves.   In such systems the pulses occur at an inaudible repetition rate, or at such a rate that the pulse frequency can be separated from the modulating signal by suitable filters.   The modulating voltage can be made to vary:

1. The height of the pulse.
2. The duration of the pulse.
3. The repetition rate.
4. The timing of the pulse with respect to a standard marker pulse.

Pulse-time modulation uses the fourth method.   In this system, a number of messages can be transmitted on a single carrier frequency.   Figure 7 shows the pulse arrangement for a five-channel system.   Each frame contains one marker pulse and five channel pulses (one for each communica-

[1] GRIEG, D. D., and A. M. LEVINE, Pulse-Time Modulated Multiplex Radio Relay System, *Elec. Commun.*, vol. 17, pp. 1061–1066; December, 1946.

[2] LACY, R. E., Two Multichannel Microwave Relay Equipments, *Proc. I.R.E.*, vol. 35, pp. 65–70; January, 1947.

tion channel). The modulating voltage of any one channel varies the timing of the corresponding pulse with respect to the marker pulse. A positive modulating voltage produces an advance in pulse time, whereas a negative modulating voltage retards the pulse. The time displacement of a pulse from its mean position is proportional to the amplitude of the modulating voltage, while the number of cycles of deviation per second is equal to the frequency of the modulating signal.

A typical system uses a marker pulse which is 4 microseconds long and channel pulses which are approximately 1 microsecond long. A 15-microsecond time interval is reserved for each pulse, which allows for a time

FIG. 7.—Pulses of a five-channel pulse-time modulation system.

variation of ±6 microseconds with a 2-microsecond interval between adjacent channels. The recurrence rate for the entire frame is 8,000 per second. The audible signal produced by the frame recurrence rate is removed in the output by low-pass filters.

Despite the fact that a large number of messages can be transmitted on a single carrier frequency, it appears that pulse-time modulation requires more frequency space than either amplitude or frequency modulation. This is due to the high harmonic content of the pulses. Pulse-time modulation, however, may have the advantage of a higher signal-to-noise ratio, particularly in systems involving a chain of relay stations. In such systems, the total distortion is the cumulative distortion of all of the relay stations in the relay link. In pulse-time systems, the relay stations contain multivibrator circuits which are triggered by the incoming pulse. In order to introduce distortion, a relay station would have to alter the timing of a pulse. This can easily be guarded against. In general, the relay stations of pulse-time systems can be much simpler than those of amplitude- or frequency-modulation systems, since a large number of channels can be amplified by a single chain of multivibrators.

CHAPTER 13

MAXWELL'S EQUATIONS

The experimental and theoretical researches of Coulomb, Ampère, Faraday, and others during the last part of the eighteenth century and first of the nineteenth century laid the foundation for an understanding of the basic principles of electric and magnetic phenomena. Coulomb experimented with the forces of attraction and repulsion between charges, Ampère demonstrated that magnetic effects are produced by an electric current, and Faraday showed that an electromotive force is induced in a conductor moving through a magnetic field.

Previously, Newton had explained gravitational forces in terms of an "action-at-a-distance" philosophy which had at first appeared incredulous but later gained widespread recognition. It was quite logical, therefore, for the early experimenters to accept this point of view as an explanation of the electric and magnetic forces which they experienced. However, Faraday's discovery that the capacitance of a condenser is dependent upon the nature of the dielectric, led him to suspect that an electric field exists in the dielectric of the condenser. Since the new field theory was in contradiction to the deeply rooted action-at-a-distance theory, a lively controversy ensued.

Maxwell, in a series of brilliant mathematical contributions, skillfully welded the electromagnetic-field concepts into a unified pattern. His introduction of the displacement current and the assumption that this produces a magnetic field led him to the prediction of electromagnetic wave phenomenon. Since his theoretical velocity of propagation proved to be very nearly equal to the velocity of light, Maxwell concluded that light itself is an electromagnetic-wave phenomenon. He visualized an all-pervading ether in which the electromagnetic waves propagate as a disturbance. Although the ether concept has since been discredited, Maxwell's fundamental concepts of electromagnetic-field theory remain as the foundation of our present-day concepts.

13.01. Fundamental Laws.—The four laws commonly referred to as Maxwell's equations include Faraday's law of induced electromotive force, Ampère's circuital law, Gauss's law for the electric field, and Gauss's law for the magnetic field. In rationalized mks units, these are

$$\oint \bar{E} \cdot d\bar{l} = -\frac{\partial \phi}{\partial t} \qquad (1)$$

$$\oint \bar{H} \cdot d\bar{l} = i \tag{2}$$

$$\oint_s \bar{D} \cdot d\bar{s} = q \tag{3}$$

$$\oint_s \bar{B} \cdot d\bar{s} = 0 \tag{4}$$

with the supplementary relationships

$$\bar{D} = \epsilon \bar{E} \tag{5}$$

$$\bar{B} = \mu \bar{H} \tag{6}$$

Faraday's law states that the emf induced in a closed path, such as that shown in Fig. 1, is equal to the rate of change of magnetic flux linking the path. No restriction is imposed as to the nature of the medium. Thus,
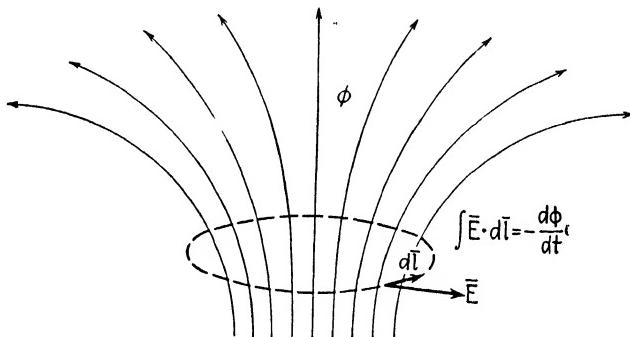
Fig. 1.—An illustration of Faraday's induced emf law.

the path over which the electric intensity is integrated may be in a conductor, in a dielectric, or in free space.

Ampère's circuital law, Eq. (2), states that the line integral of magnetic intensity around a closed path is equal to the current linking the path. The line integral $\oint \bar{H} \cdot d\bar{l}$ is known as the magnetomotive force. Ampère's law is illustrated in Fig. 2, in which the dotted-line path indicates the path of integration and i is the current flowing in the conductor. Maxwell showed that the current i must include not only the conduction (or convection) current, but also the displacement current. The displacement current is necessary in order to explain electromagnetic-wave phenomenon in free space or in dielectrics where the conduction and convection currents may be negligible.

Equation (3) is Gauss's law which was discussed in Sec. 2.04. This states that the net outward electric flux through any closed surface is equal to the charge enclosed by the surface. In electrostatic fields the
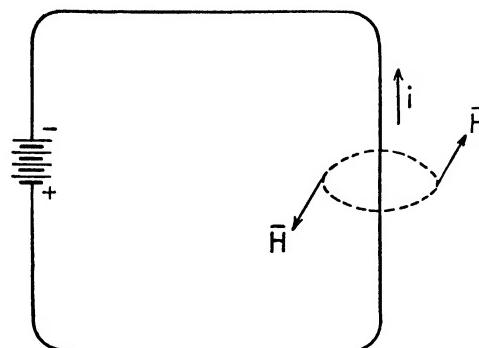
FIG. 2.—An illustration of Ampère's circuital law.

electric flux lines begin and end on charges. In time-varying fields, how-ever, the electric flux lines can exist as closed loops.

Gauss's law for the magnetic field, Eq. (4), states that the net outward magnetic flux through any closed surface is zero. This is equivalent to stating that magnetic flux lines are always continuous, and thus form closed loops.

In rationalized mks units, E is in volts per meter, D in coulombs per square meter, q in coulombs, and i in amperes. Among the magnetic-field quantities, we have H in ampere turns per meter, ϕ in webers, and B in webers per square meter. The permittivity is given by $\epsilon = \epsilon_0 \epsilon_r$, where $\epsilon_0 = 8.85 \times 10^{-12}$ farad per meter is the permittivity of free space and ϵ_r is the relative permittivity (or dielectric constant). Similarly, we have $\mu = \mu_0 \mu_r$, where $\mu_0 = 4\pi \times 10^{-7}$ henry per meter is the permeability of free space and μ_r is the relative permeability. The relative perme-ability of nonmagnetic materials may be taken as unity.

13.02. The Curl.—Maxwell's equations become exceedingly powerful tools when ex-pressed in differential-equation form. The

FIG. 3.—An illustration of Gauss's law.

differential-equation forms of Faraday's and Ampère's laws are stated in terms of the curl of a vector. Let us therefore derive an expression for the curl, as applied to Faraday's law. In Eq. (13.01-1), the electric in-

tensity $\overline{E}$ may be interpreted as the force on unit positive charge placed in the field, and the integral $\oint \overline{E} \cdot d\overline{l}$ then becomes the work done by the field in moving unit positive charge around a closed path (the path of integration).
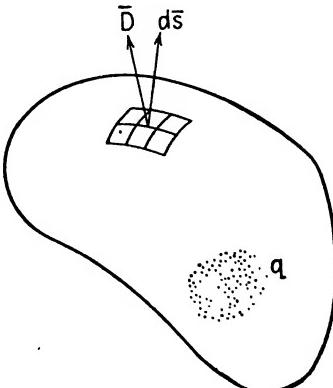
Now consider the work done in carrying unit charge around the perimeter of the differential area shown in Fig. 4. The differential area is assumed to be oriented parallel to the yz plane. Let dw be the work done in carrying unit charge around the closed path $abcda$. We shall define $\mathrm{curl}_x \overline{E}$ by the relationship
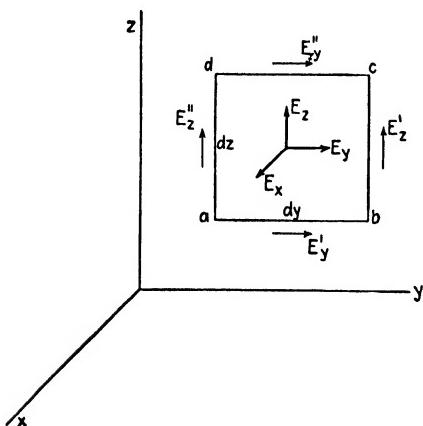
FIG. 4.—The derivation of $\mathrm{curl}_x \overline{E}$.

$$\mathrm{curl}_x \overline{E} = \lim_{dy\,dz \to 0} \frac{dw}{dy\,dz}\,\overline{\imath} \quad (1)$$

Thus, $\mathrm{curl}_x \overline{E}$ is the work done in carrying unit charge around the perimeter of the differential area, divided by the area. Dividing by the area gives us the work per unit area. However, it should be noted that the curl is computed for a vanishingly small area and is therefore a point function. $\mathrm{curl}_x \overline{E}$ is a vector, having a direction perpendicular to the area $dy\,dz$.

Let $\overline{E} = E_x\overline{\imath} + E_y\overline{\jmath} + E_z\overline{k}$ be the electric intensity at the center of the differential area. The electric-intensity components at the sides of the differential area are then

$$E'_y = E_y + \frac{\partial E_y}{\partial z}\left(-\frac{dz}{2}\right)$$

$$E'_z = E_z + \frac{\partial E_z}{\partial y}\left(\frac{dy}{2}\right)$$

$$E''_y = E_y + \frac{\partial E_y}{\partial z}\left(\frac{dz}{2}\right) \quad (2)$$

$$E''_z = E_z + \frac{\partial E_z}{\partial y}\left(-\frac{dy}{2}\right)$$

The work done in moving unit charge around the closed path may be obtained by multiplying the electric intensity (or force on unit charge) times distance of travel. Thus, referring to Fig. 4, the first and second

expressions in Eq. (2) are multiplied by dy and dz, respectively, while the third and fourth are multiplied by $-dy$ and $-dz$, respectively. Adding these four terms to obtain the work done in carrying unit charge around the closed path, and simplifying, we obtain

$$dw = \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right) dy\, dz \tag{3}$$

Substitution of Eq. (3) into (1) gives, for $\text{curl}_x \bar{E}$

$$\text{curl}_x \bar{E} = \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right) \bar{\imath} \tag{4}$$

The differential area could also have been oriented parallel to the xz and the xy planes to obtain two other curl components,

$$\text{curl}_y \bar{E} = \left(\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}\right) \bar{\jmath} \qquad \text{curl}_z \bar{E} = \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}\right) \bar{k} \tag{5}$$

Equations (4) and (5) give the three components of a resultant vector quantity symbolized by $\text{curl}\,\bar{E}$. Adding the components, we obtain

$$\text{curl}\,\bar{E} = \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right) \bar{\imath} + \left(\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}\right) \bar{\jmath} + \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}\right) \bar{k} \tag{6}$$

If we take the differential operator ∇ given by Eq. (2.03-2) and perform the operation $\nabla \times \bar{E}$, using the cross-product rule outlined in Sec. 2.01, the resulting expression gives the terms on the right-hand side of Eq. (6). The curl may therefore be written in the determinental form

$$\text{curl}\,\bar{E} = \nabla \times \bar{E} = \begin{vmatrix} \bar{\imath} & \bar{\jmath} & \bar{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ E_x & E_y & E_z \end{vmatrix} \tag{7}$$

Expressions for the curl in cylindrical and spherical coordinates are given in Appendix III. In expanding these determinants, the second-row differentials operate on the third-row quantities.

Thus far we have been concerned with the derivation of a mathematical expression for the curl. Let us now use Faraday's law to relate $\text{curl}\,\bar{E}$ to the magnetic-field quantities. We first write Eq. (1) in the form

$$dw\,\bar{\imath} = \text{curl}_x \bar{E}\, dy\, dz \tag{8}$$

The work dw is the value of the line integral $\oint \bar{E} \cdot d\bar{l}$ around the differential area. According to Faraday's law, Eq. (13.01-1), this must be equal to

the negative time rate of change of magnetic flux through the differential area.   The magnetic flux through differential area is $\phi = B_x \, dy \, dz$, where B_x is the x-directed component of magnetic flux density.   Hence, Faraday's law yields

$$\mathrm{curl}_x \, \bar{E} \, dy \, dz \,=\, -\frac{\partial B_x}{\partial t} \, dy \, dz \, \bar{\imath}$$

$$\mathrm{curl}_x \, \bar{E} \,=\, -\frac{\partial B_x}{\partial t} \, \bar{\imath}$$

In a similar manner, we may obtain $\mathrm{curl}_y \, \bar{E} = (-\partial B_y/\partial t)\bar{\jmath}$ and $\mathrm{curl}_z \, \bar{E} = (-\partial B_z/\partial t)\bar{k}$.   The whole curl, obtained by adding the component curls, is therefore

$$\mathrm{curl} \, \bar{E} \,=\, \nabla \times \bar{E} \,=\, -\frac{\partial \bar{B}}{\partial t} \tag{9}$$

where

$$\bar{B} = B_x \bar{\imath} + B_y \bar{\jmath} + B_z \bar{k}$$

Equation (9) is the differential equation form of Faraday's induced-emf law.   Each component of the curl was defined as the work done by the field in moving unit charge around the perimeter of a differential area, divided by the area—or briefly, the work per unit area.   The three orthogonal orientations of the differential area at any one point in space yield the three curl components, the whole curl being the vector sum of the three components.   There is always a particular orientation of the differential area, such that the curl computed for this area gives the whole curl without bothering with the components.   This occurs when the differential area is oriented in a direction normal to the direction of the magnetic flux-density vector $\bar{B}$.

13.03. Useful Vector-analysis Relationships.—The divergence theorem and Stokes's theorem are two vector-analysis relationships which facilitate the derivation of the differential equation form of Maxwell's equations. Let us therefore consider these relationships.

The divergence theorem may be written

$$\oint_s \bar{A} \cdot d\bar{s} \,=\, \int_\tau (\nabla \cdot \bar{A}) \, d\tau \tag{1}$$

This states that the surface integral of a vector $\bar{A}$ is equal to the volume integral of the divergence of $\bar{A}$.

In order to visualize a physical interpretation of the divergence theorem, let us assume that the vector $\bar{A}$ is a flux density.   The quantity $\int_s \bar{A} \cdot d\bar{s}$ is then the net outward flux through the given surface.   In Sec. 2.05,

the divergence of $\bar{A}$ was defined as the net outward flux through a differential volume, divided by the volume.   Now consider a volume of any size, which we divide into a very large number of differential volume elements.   In computing the divergence for the differential elements, we find that part of the flux flows out of one element and into an adjacent element.   This flux produces equal and opposite contributions to the divergences of the respective elements.   In summing up all of the divergences, therefore, we find that the flux which is common to adjacent elements cancels.   The only flux which does not cancel is that through the outer surface of the volume.   Hence, the summation of the divergences throughout the entire volume gives the net outward flux through the bounding surface.

Stokes's theorem may be expressed as

$$\oint \bar{A} \cdot d\bar{l} = \int_s (\nabla \times \bar{A}) \cdot d\bar{s} \qquad (2)$$

This states that the line integral of the vector $\bar{A}$ around a closed path is equal to the surface integral of the curl over any surface bounded by the closed path.

FIG.   5.—An   illustration   of Stokes's theorem.

To simplify the physical interpretation, let us replace the vector $\bar{A}$ by the electric intensity $\bar{E}$.   The integral $\oint \bar{E} \cdot d\bar{l}$ may be interpreted as the work done by the field in moving unit charge around a closed path.   Referring to Fig. 5, let us divide the area shown into a large number of differential area elements.   The curl of any element may be considered as the work done in moving unit charge around the differential element, divided by the area of the element.   We note that the differential elements have interior sides which are common to two adjacent elements and exterior sides bordering on the perimeter.   In evaluating the work of two adjacent elements, the common side is traversed in opposite directions.   When these two works are added, the net work done in carrying the unit charge along the common side is zero.   Thus, in summing the curls for the entire area, we find that the work done along the paths which are common to two differential elements cancels.   The only sides which contribute any net work are those bordering on the perimeter.   The work done in carrying the unit charge along these sides is $\oint \bar{E} \cdot d\bar{l}$.   Hence, the summation of the infinitesimal contributions over the given area is equal to the line integral over the perimeter of the area.

Two additional vector analysis identities will aid in the physical and mathematical interpretation of the field equations. The first of these states that the curl of a gradient is always zero, that is

$$\nabla \times (\nabla V) = 0 \tag{3}$$

This relationship may be used to point out certain limitations in commonly used field equations. In Eq. (2.03-3), the electric intensity was expressed as $\bar{E} = -\nabla V$. If this is inserted into Eq. (3), there results $\nabla \times \bar{E} = 0$. This is not in agreement with Eq. (13.02-9) except for the special case where $\partial \bar{B}/\partial t = 0$. Since Eqs. (3) and (13.02-9) are always valid, we conclude that the relationship $\bar{E} = -\nabla V$, as well as its counterpart $V = -\int \bar{E} \cdot d\bar{l}$, and Poisson's equation, which was derived from the gradient relationship, are valid only when $\partial \bar{B}/\partial t = 0$, *i.e.*, in stationary fields.

A field in which the curl is everywhere zero is known as an *irrotational field*. The foregoing discussion may be generalized by the following vector-analysis theorem: *A vector field may be expressed as the gradient of a scalar potential only if the field is irrotational.*

The second useful vector-analysis identity states that the divergence of the curl of a vector quantity is always zero, or

$$\nabla \cdot (\nabla \times \bar{A}) = 0 \tag{4}$$

A useful corollary of this identity states that if the divergence of a vector field is everywhere zero, the vector field may be expressed as the curl of another vector. A vector field which has zero divergence is known as a *solenoidal field*. We shall find that magnetic fields are solenoidal fields.

13.04. Maxwell's Equations in Differential-equation Form.—Let us now use the divergence theorem and Stokes's theorem to obtain the differential-equation form of Maxwell's equations. Returning to Faraday's law, Eq. (13.01-1), and applying Stokes's theorem, Eq. (13.03-2), we obtain

$$\oint \bar{E} \cdot d\bar{l} = \int_s (\nabla \times \bar{E}) \cdot d\bar{s} = -\frac{\partial \phi}{\partial t} \tag{1}$$

The magnetic flux may be expressed as the surface integral of flux density, thus $\phi = \int_s \bar{B} \cdot d\bar{s}$. Inserting this into Eq. (1) and interchanging the order of differentiation and integration, we obtain

$$\int_s (\nabla \times \bar{E}) \cdot d\bar{s} = -\int_s \left(\frac{\partial \bar{B}}{\partial t}\right) \cdot d\bar{s} \tag{2}$$

Both sides of Eq. (2) are integrated over the same surface area, hence we may equate the integrands, yielding the curl equation

$$\nabla \times \bar{E} = -\frac{\partial \bar{B}}{\partial t} \tag{3}$$

This is identical to Eq. (13.02-9).

The differential equation form of Ampère's law may be obtained by a similar procedure. Combining Eqs. (13.03-2), written in terms of magnetic intensity, and Eq. (13.01-2), we get

$$\oint \bar{H} \cdot d\bar{l} = \int_s (\nabla \times \bar{H}) \cdot d\bar{s} = i \tag{4}$$

The current may be expressed as the surface integral of current density, thus $i = \int_s \bar{J} \cdot d\bar{s}$. Inserting this into Eq. (4), we have

$$\int_s (\nabla \times \bar{H}) \cdot d\bar{s} = \int_s \bar{J} \cdot d\bar{s} \tag{5}$$

Again the integration of both sides of the equation is over the same surface area; hence the integrands may be equated to obtain the curl equation for the magnetic field,

$$\nabla \times \bar{H} = \bar{J} \tag{6}$$

The current density J includes the conduction current density and displacement current density as expressed by Eq. (3.02-5). When this is substituted into Eq. (6), there results

$$\nabla \times \bar{H} = \bar{J}_c + \frac{\partial \bar{D}}{\partial t} \tag{7}$$

The differential equation form of Gauss's law may be derived by the use of the divergence theorem. Combining Eqs. (13.01-3) and (13.03-1), we obtain

$$\oint_s \bar{D} \cdot d\bar{s} = \int_\tau (\nabla \cdot \bar{D}) \, d\tau = q \tag{8}$$

Electric charge can be expressed as the volume integral of charge density, thus $q = \int_\tau q_\tau \, d\tau$. Inserting this into Eq. (8), we obtain

$$\int_\tau (\nabla \cdot \bar{D}) \, d\tau = \int_\tau q_\tau \, d\tau \tag{9}$$

Both sides of this equation are integrated over the same volume, hence, upon equating integrands, we obtain the divergence equation

$$\nabla \cdot \bar{D} = q_\tau \tag{10}$$

A similar procedure applied to Eq. (13.01-4) yields the divergence equation for the magnetic field,

$$\nabla \cdot \bar{B} = 0 \qquad (11)$$

The foregoing relationships are summarized in Table 3 at the end of this chapter.

13.05. The Wave Equations.—The curl equations (13.04-3) and (13.04-7) contain electric and magnetic quantities in each equation. These two equations may be solved as simultaneous equations to obtain two explicit equations, one containing the electric intensity and the other containing the magnetic intensity. These are known as the *wave equations*. They are more convenient than the curl equations for the solution of many types of electromagnetic-field problems.

Before proceeding with the derivation of the wave equations, let us state a useful vector-analysis identity for the curl of the curl of a vector quantity. This may be written

$$\nabla \times \nabla \times \bar{A} = -\nabla^2 \bar{A} + \nabla(\nabla \cdot \bar{A}) \qquad (1)$$

Now take the curl of both sides of Eq. (13.04-3) thus,

$$\nabla \times \nabla \times \bar{E} = -\mu \nabla \times \frac{\partial \bar{H}}{\partial t} = -\mu \frac{\partial}{\partial t}(\nabla \times \bar{H}) \qquad (2)$$

The step in going from the second to the third form of Eq. (2) amounts to interchanging of the order of differentiation.

Now substitute Eq. (1) for the left-hand side of Eq. (2), and Eq. (13.04-7) for $\nabla \times \bar{H}$ on the right-hand side. In place of $\bar{J}_c$, we write $\bar{J}_c = \sigma \bar{E}$, thus obtaining

$$-\nabla^2 \bar{E} + \nabla(\nabla \cdot \bar{E}) = -\mu \left[\sigma \frac{\partial \bar{E}}{\partial t} + \epsilon \frac{\partial^2 \bar{E}}{\partial t^2} \right] \qquad (3)$$

In most applications, the space-charge density is zero; hence Eq. (13.04-10) gives $\nabla \cdot \bar{E} = 0$. Equation (3) may then be written

$$\nabla^2 \bar{E} = \mu \left(\sigma \frac{\partial \bar{E}}{\partial t} + \epsilon \frac{\partial^2 \bar{E}}{\partial t^2} \right) \qquad (4)$$

This is the wave equation for the electric field. A similar expression may be derived for the magnetic field by taking the curl of both sides of Eq. (13.04-7) and substituting Eqs. (1), (13.04-3), and (13.04-11). This yields the wave equation for the magnetic field,

$$\nabla^2 \bar{H} = \mu \left(\sigma \frac{\partial \bar{H}}{\partial t} + \epsilon \frac{\partial^2 \bar{H}}{\partial t^2} \right) \qquad (5)$$

The use of the wave equations will be considered in the following chapter

13.06. Fields with Sinusoidal Time Variation.—The principal electromagnetic-field equations can be simplified for fields with sinusoidal time variation. We shall assume a time variation of the form $\epsilon^{j\omega t}$. The time derivatives may then be written $\partial \bar{E}/\partial t = j\omega \bar{E}$ and $\partial^2 \bar{E}/\partial t^2 = -\omega^2 \bar{E}$. Making these substitutions in Eqs. (13.04-3 and 7) and (13.05-4 and 5), we obtain the curl equations and wave equations

$$\nabla \times \bar{E} = -j\omega\mu\bar{H} \tag{1}$$

$$\nabla \times \bar{H} = (\sigma + j\omega\epsilon)\bar{E} \tag{2}$$

$$\nabla^2 \bar{E} = j\omega\mu(\sigma + j\omega\epsilon)\bar{E} = \gamma^2 E \tag{3}$$

$$\nabla^2 \bar{H} = j\omega\mu(\sigma + j\omega\epsilon)\bar{H} = \gamma^2 \bar{H} \tag{4}$$

where

$$\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)} = \alpha + j\beta \tag{5}$$

The quantity γ is a property of the medium known as the *intrinsic propagation constant*. It is analogous to the propagation constant of the transmission line. In general, γ is complex, its real part being the attenuation constant α and imaginary part the phase constant β.

13.07. Power Flow and Poynting's Vector.—The energy density stored in an electromagnetic field can be represented by [1]

$$w = \tfrac{1}{2}(\epsilon E^2 + \mu H^2) \tag{1}$$

where $\tfrac{1}{2}\epsilon E^2$ is the energy density of the electric field and $\tfrac{1}{2}\mu H^2$ is the energy density of the magnetic field. In mks units the energy density is in joules per cubic meter.

Let us now consider the concept of Poynting's vector, which we shall have frequent occasion to use in evaluating the power flow. We start by taking the divergence of $\bar{E} \times \bar{H}$. A vector-analysis identity yields

$$\nabla \cdot (\bar{E} \times \bar{H}) = \bar{H} \cdot (\nabla \times \bar{E}) - \bar{E} \cdot (\nabla \times \bar{H}) \tag{2}$$

Substitution of Eqs. (13.04-3 and 7) for $\nabla \times \bar{E}$ and $\nabla \times \bar{H}$ in Eq. (2) yields

$$\nabla \cdot (\bar{E} \times \bar{H}) = -\mu\bar{H} \cdot \frac{\partial \bar{H}}{\partial t} - \bar{E} \cdot \left(\sigma\bar{E} + \epsilon \frac{\partial \bar{E}}{\partial t}\right) \tag{3}$$

The dot product of a vector with itself is equal to the scalar quantity squared, that is, $\bar{E} \cdot \bar{E} = E^2$. Also, we may write

$$\bar{E} \cdot \frac{\partial \bar{E}}{\partial t} = \frac{1}{2} \frac{\partial(\bar{E} \cdot \bar{E})}{\partial t} = \frac{1}{2} \frac{\partial(E)^2}{\partial t}$$

[1] FRANK, N. H., "Introduction to Electricity and Optics," pp. 52–134, McGraw-Hill Book Company, Inc., New York, 1940. Equations given in this reference are in unrationalized units and therefore differ from the above equations by a factor of 4π.

A similar expression may be written for the magnetic intensity.   Inserting these into Eq. (3), we obtain

$$-\nabla \cdot (\bar{E} \times \bar{H}) = \frac{\partial}{\partial t} \frac{1}{2}(\epsilon E^2 + \mu H^2) + \sigma E^2 \tag{4}$$
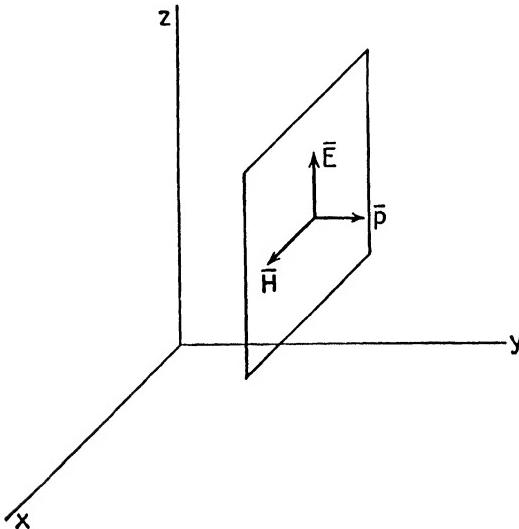
FIG. 6.—Poynting's vector is perpendicular to $\bar{E}$ and $\bar{H}$.

The vector quantity $\bar{E} \times \bar{H}$ in the first term of Eq. (4) is known as the *Poynting vector*, thus

$$\bar{\mathcal{P}} = \bar{E} \times \bar{H} \tag{5}$$

Poynting's vector may be interpreted as the power-density flow per unit area.   It is represented by a vector which is perpendicular to the $\bar{E}$ and $\bar{H}$ vectors.
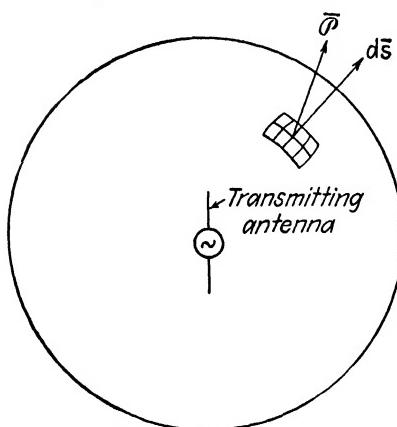
FIG. 7.—Application of Poynting's vector.

Divergence is the net outward flow per unit volume, hence $-\nabla \cdot (\bar{E} \times \bar{H})$ represents the net *inward flow* of power per unit volume.   Part of this power contributes to an increase in the energy storage in the electric and magnetic fields, while part of it is lost owing to imperfect conductivity of the medium.   The quantity $\partial/\partial t\, [\frac{1}{2}(\epsilon E^2 + \mu H^2)]$ on the right-hand side of Eq. (4) is the time rate of change of energy storage in the field.   There-

fore, this term represents that portion of the power which contributes to an increase in energy storage. The second term on the right-hand side represents the power loss due to imperfect conductivity. Thus, Eq. (4) states that the net inward power flow per unit volume is equal to the rate of change of energy density stored in the field plus the loss per unit volume due to imperfect conductivity.

We may obtain an expression for power flow through any surface by integrating the left-hand side of Eq. (4) over the volume enclosed by the surface and applying the divergence theorem. Representing the total power flow by p, we have

$$p = \int_\tau \nabla \cdot (\bar{E} \times \bar{H}) \, d\tau = \oint_s (\bar{E} \times \bar{H}) \cdot d\bar{s} \tag{6}$$

Hence the total power flow through any closed surface is equal to the surface integral of the normal component of Poynting's vector over that surface. As an example, the power radiated by the antenna of Fig. 7 may be computed by integrating the normal component of Poynting's vector over the surface enclosing the antenna.

13.08. Boundary Conditions.—In order for a given electromagnetic field distribution to exist, it must: (1) be a solution of Maxwell's equations and (2) satisfy certain boundary conditions for the given physical system. Let us therefore consider these boundary conditions.

First, consider the tangential components of electric intensity on either side of a geometrical surface which is the interface between two different mediums as shown in Fig. 8. Assume that a unit charge is carried a short distance Δz parallel to the boundary at the surface in medium 1 as shown in Fig. 8, then an equal distance along the surface in medium 2, to return to the starting point. The work done in carrying the charge around the closed path is

$$\oint \bar{E} \cdot d\bar{l} = (E_{t2} - E_{t1}) \, \Delta z \tag{1}$$

From Faraday's law, we have $\oint \bar{E} \cdot d\bar{l} = -\partial \phi / \partial t$. Since the two paths are assumed to be an infinitesimal distance apart, the magnetic flux linking the path is vanishingly small, and $\oint \bar{E} \cdot d\bar{l} = 0$. Equation (1) therefore yields

$$E_{t1} = E_{t2} \tag{2}$$

or, the tangential components of electric intensity are continuous across the boundary.

Gauss's law enables us to establish a relationship between the normal components of electric flux density at the boundary. Consider a small

surface enclosing an incremental boundary area Δs, as shown at the bottom of Fig. 8.   If the surface charge density is q_s, the charge enclosed is $q_s \, \Delta s$. Gauss's law then yields

$$\oint_s \bar{D} \cdot d\bar{s} = (D_{n2} - D_{n1}) \, \Delta s = q_s \, \Delta s$$

or

$$D_{n2} - D_{n1} = q_s \tag{3}$$

Therefore, the discontinuity in the normal component of electric flux density is equal to the surface charge density.   In most problems with which we
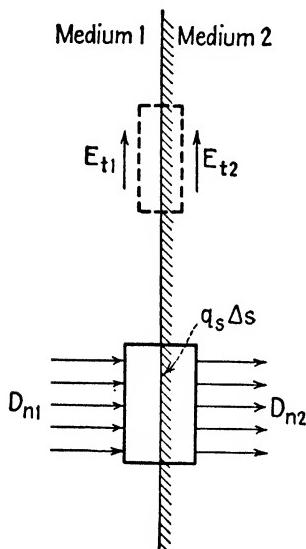
FIG. 8.—Boundary conditions for the electric field
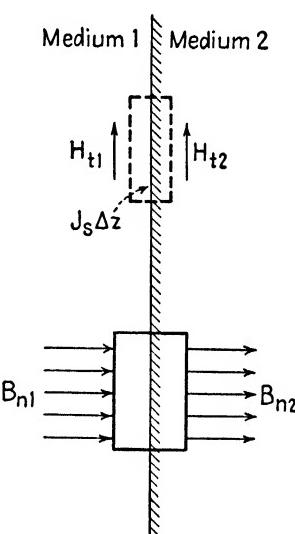
FIG. 9.—Boundary conditions for the magnetic field.

will be concerned, we may assume that $q_s = 0$; hence $D_{n2} = D_{n1}$ or $E_{n1}/E_{n2} = \epsilon_2/\epsilon_1$.

To obtain a relationship for the tangential components of the magnetic field, we use Ampère's law.   The magnetic intensity $\bar{H}$ may be interpreted as the force on a fictitious unit magnetic pole, and thus, $\oint \bar{H} \cdot d\bar{l}$ is the work done in carrying the unit pole around a closed path.   If such a pole is carried a distance Δz parallel to the boundary at the surface of medium 1 and an equal distance in the opposite direction in medium 2 to return to the starting point, the work done becomes

$$\oint \bar{H} \cdot d\bar{l} = (H_{t1} - H_{t2}) \, \Delta z \tag{4}$$

Applying Ampère's law, we have $\oint \bar{H} \cdot d\bar{l} = i$, where i is the current flowing through the area enclosed by the line integral. Since H_{t1} and H_{t2} are assumed to be an infinitesimal distance apart, the area is vanishingly small and, in the limit, we have $\oint \bar{H} \cdot d\bar{l} = 0$. Thus, Eq. (4) yields $H_{t1} = H_{t2}$. There is one important exception to this statement. In the theoretical case of a perfect conductor ($\sigma = \infty$), an infinitely thin current sheet can flow on the geometrical surface of the conductor. This current flows in a direction perpendicular to H_t. The current per unit length of surface will be referred to as the *surface current density*, J_s. Thus, the surface current flowing through the infinitesimal area between H_{t1} and H_{t2} of height Δz is $J_s \, \Delta z$, and Eq. (4) yields

$$H_{t1} - H_{t2} = J_s \tag{5}$$

The discontinuity in the tangential component of magnetic intensity is therefore equal to the surface current density. In all cases except that of a perfect conductor, the surface current density is vanishingly small, and we have $H_{t1} = H_{t2}$.

Applying Gauss's law for the magnetic field in a manner similar to that for the electric field, we obtain

$$B_{n1} = B_{n2} \tag{6}$$

Thus, the boundary conditions may be summarized as follows:

1. The tangential components of electric intensity are continuous across the boundary.

2. The normal components of electric flux density differ by an amount equal to the surface charge density.

3. The tangential components of magnetic intensity differ by an amount equal to the surface current density. The surface current density is vanishingly small and may be assumed to be zero in all cases except that of a perfect conductor.

4. The normal components of magnetic flux density are equal.

Tables 3 and 4 summarize a number of the more important electromagnetic field relationships given in this chapter and in Chap. 14.

Table 3.—Principal Electromagnetic Field Equations

	Differential-equation form	
Integral form	General time function	Sinusoidal time function
Faraday's law .... $\oint \bar{E} \cdot d\bar{l} = -\dfrac{\partial \phi}{\partial t}$ (13.01–1)	$\nabla \times \bar{E} = -\mu \dfrac{\partial \bar{H}}{\partial t}$ (13.04–3)	$\nabla \times \bar{E} = -j\omega\mu \bar{H}$ (13.06–1)
Ampère's law .... $\oint \bar{H} \cdot d\bar{l} = i$ (13.01–2)	$\nabla \times \bar{H} = J_c + \epsilon \dfrac{\partial \bar{E}}{\partial t}$ (13.04–7)	$\nabla \times \bar{H} = (\sigma + j\omega\epsilon)\bar{E}$ (13.06–2)
Gauss's law (electric field) .... $\oint \bar{D} \cdot d\bar{s} = q$ (13.01–3)	$\nabla \cdot \bar{E} = \dfrac{q_r}{\epsilon}$ (13.04–10)	$\nabla \cdot \bar{E} = \dfrac{q_r}{\epsilon}$ (13.04–10)
Gauss's law (magnetic field) .... $\oint \bar{B} \cdot d\bar{s} = 0$ (13.01–4)	$\nabla \cdot \bar{H} = 0$ (13.04–11)	$\nabla \cdot \bar{H} = 0$ (13.04–11)
Wave equation (electric intensity)	$\nabla^2 \bar{E} = \mu \left[\sigma \dfrac{\partial \bar{E}}{\partial t} + \epsilon \dfrac{\partial^2 \bar{E}}{\partial t^2} \right]$ (13.05–4)	$\nabla^2 \bar{E} = \gamma^2 \bar{E}$ (13.06–3)
Wave equation (magnetic intensity)	$\nabla^2 \bar{H} = \mu \left[\sigma \dfrac{\partial \bar{H}}{\partial t} + \epsilon \dfrac{\partial^2 \bar{H}}{\partial t^2} \right]$ (13.05–5)	$\nabla^2 \bar{H} = \gamma^2 \bar{H}$ (13.06–4)
Conservation of charge .... $\displaystyle\int J_c \cdot d\bar{s} = -\dfrac{dq}{dt}$	$\nabla \cdot J_c = -\dfrac{\partial q_r}{\partial t}$ (3.02–7)	$\nabla \cdot J_c = -j\omega q_r$

TABLE 4.—ADDITIONAL RELATIONSHIPS

Electric flux density.................	$\bar{D} = \epsilon \bar{E}$	(13.01–5)
Magnetic flux density.................	$\bar{B} = \mu \bar{H}$	(13.01–6)
Current density.................	$J = J_c + \epsilon \dfrac{\partial \bar{E}}{\partial t}$	(3.02–5)
Convection-current density.................	$J_c = q_r \bar{v}$	(3.01–1)
Conduction-current density.................	$J_c = \sigma \bar{E}$	(3.01–2)
Force on a charge.................	$\bar{f} = q(\bar{E} + \bar{v} \times \bar{B})$	(7.04–1)
Propagation constant........ (sinusoidal time variation)	$\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)}$	(13.06–5)
Intrinsic impedance........ (sinusoidal time variation)	$\eta = j\dfrac{\omega\mu}{\gamma} = \dfrac{\gamma}{\sigma + j\omega\epsilon} = \sqrt{\dfrac{j\omega\mu}{\sigma + j\omega\epsilon}}$	(14.02–7)
Time-average power density........ (perfect dielectric)	$\mathscr{P}_{av} = \dfrac{\eta\,|H|^2}{2} = \dfrac{|E|^2}{2\eta}$	(14.04–3)
Time-average power density........ (good conductor)	$\mathscr{P}_{av} = \dfrac{|H|^2}{2}\sqrt{\dfrac{\omega\mu}{2\sigma}}$	(14.04–4)
Depth of penetration.................	$\delta = \dfrac{1}{\alpha} = \sqrt{\dfrac{2}{\omega\mu\sigma}}$	(14.07–1)
Skin-effect resistance.................	$\mathscr{R}_s = \dfrac{1}{\sigma\delta} = \sqrt{\dfrac{\omega\mu}{2\sigma}}$	(14.07–4)

PROBLEMS

1. Given a vector $\bar{A} = (x^2 - y^2)\bar{\imath} + 2xy\bar{\jmath}$
   - (a) Evaluate $\nabla \times \bar{A}$.
   - (b) Evaluate the surface integral $\int_s (\nabla \times \bar{A}) \cdot d\bar{s}$ over a surface in the xy plane bounded by the four lines $x = 0$, $y = b$, $x = a$, and $y = 0$.
   - (c) Evaluate $\oint \bar{A} \cdot d\bar{l}$ around the perimeter of the rectangle.
   - (d) Show that Stokes's theorem $\int_s (\nabla \times \bar{A}) \cdot d\bar{s} = \oint \bar{A} \cdot d\bar{l}$ applies.

2. Prove the relationship $\nabla \times (\nabla V) = 0$ using rectangular coordinates.

3. Prove that $\nabla \times (\nabla \times \bar{A}) = -\nabla^2 \bar{A} + \nabla(\nabla \cdot \bar{A})$ using rectangular coordinates.

4. Prove that $\nabla \cdot (\bar{A} \times \bar{B}) = \bar{B} \cdot (\nabla \times \bar{A}) - \bar{A} \cdot (\nabla \times B)$ using rectangular coordinates.

5. Derive the wave Eq. (13.05-5) for the magnetic intensity $\bar{H}$.

6. Write Maxwell's equations in differential equation form and the wave Eqs. (13.05-4 and 5) for the following special cases:
   - (a) Stationary electric and magnetic fields
   - (b) A lossless dielectric medium (with time-varying fields)
   - (c) A good conductor (with time-varying fields)

7. A long coaxial line has a d-c generator at one end and a load resistance at the other end. The conductors are assumed to have finite conductivity, but the dielectric between conductors is assumed to be lossless. On an enlarged view of the coaxial line, show by means of arrows the directions of the electric intensity, magnetic intensity, and Poynting's vector in the dielectric between conductors and in each of the conductors. Explain the mechanism of power flow along the line and of power loss in the conductors in terms of your diagram. How would you compute the power flow down the line and power loss in the conductor if the field intensities were known?

8. In deriving the wave equations, the space-charge density in the interior of the medium was assumed to be zero. The question arises as to whether or not this assumption is valid in the interior of metals where there is a plentiful supply of free charges. To investigate this, write Eq. (13.04-7) in the form $\nabla \times \bar{H} = \sigma \bar{E} + \epsilon(\partial \bar{E}/\partial t)$. Now take the divergence of both sides of this expression and use Eqs. (13.03-4) and (13.04-10) to obtain a differential equation involving the space-charge density q_τ as a function of time. The solution of this equation yields $q_\tau = e^{-(\sigma t/\epsilon) + c}$. The constant c may be evaluated by assuming that at zero time the space-charge density is $q_{\tau 0}$.

   Carry through the foregoing derivation and obtain the expression for q_τ as a function of $q_{\tau 0}$ and time. Show that the space-charge density decreases at an exponential rate which is independent of any applied fields. The *relaxation time* is the time required for q_τ to decrease to $1/e$ of its original value. Compute the relaxation time for silver, letting $\sigma = 6.14 \times 10^7$ mhos per m and $\epsilon = 8.85 \times 10^{-12}$ farad per m. Compare this relaxation time with the period of a 3,000-megacycle wave. Is the assumption of $q_\tau = 0$ justified when deriving the wave equation for metals?

9. Show that at the boundary of an imperfect conductor, the tangential magnetic intensity is equal to the current flowing through a section of conductor of unit height and infinite depth (in a direction perpendicular to the boundary surface). Explain what happens to the electric and magnetic fields in the conductor and the current and current density as the conductivity approaches infinity.

10. Show that if the boundary conditions are satisfied for either the electric intensity or the magnetic intensity, and the fields satisfy Maxwell's equations, then the boundary conditions for the other intensity are automatically satisfied.

CHAPTER 14

PROPAGATION AND REFLECTION OF PLANE WAVES

We can conceive of electromagnetic waves as being initiated by the motion of charged particles. The particle motion produces a disturbance in the field which, under proper circumstances, can take the form of an electromagnetic wave propagating outward from the source with a velocity equal to the velocity of light.

Maxwell's equations show that a time-varying electric field produces a magnetic field and, conversely, that a time-varying magnetic field produces an electric field. By virtue of this interrelationship, the electric and magnetic fields can propagate each other in the form of an electromagnetic wave. Although the wave is initiated by the motion of charged particles, once the wave starts on its journey, we may consider it to be detached from the charged particles at the source. The propagation characteristics of the wave are then determined solely by the electrical characteristics of the medium through which the wave travels.

If a small antenna is isolated in space and is radiating energy, the radiated waves are essentially spherical. A spherical wave is one in which the equiphase surfaces are concentric spheres. These equiphase spheres expand as the wave travels outward from the source. To an observer who is situated at a remote distance from the antenna, the wave would appear substantially as a uniform plane wave, since he is able to observe only a very limited portion of the wavefront. This is analogous to the observer on the surface of the earth who sees the earth's surface as a plane, since he is able to view only a small portion of the total surface.

In this chapter we shall consider the propagation characteristics of uniform plane waves in various mediums and the reflection of plane waves at boundary surfaces. We shall find that the expressions for the electric and magnetic intensities may be written in a form similar to the expressions for voltage and current on transmission lines. In fact, it is possible to carry over most of our methods of transmission-line analysis and apply them directly to problems dealing with plane-wave propagation and reflection.

14.01. Uniform Plane Waves in a Lossless Dielectric Medium.—In order to illustrate the use of Maxwell's equations, let us consider the elementary case of a uniform plane wave in a homogeneous dielectric medium.

265

The dielectric is assumed to be lossless; hence we have $\sigma = 0$. It is also assumed that the electric intensity is polarized in the x direction, as given by $\bar{E} = E_x \bar{\imath}$, and that the wave is traveling in the y direction. The magnetic intensity is perpendicular to both the electric intensity and the direction of propagation of the wave; therefore we have $\bar{H} = H_z \bar{k}$. Since a uniform plane wave has no variation of intensity in a plane normal to the direction of propagation of the wave, we have $\partial/\partial x = \partial/\partial z = 0$.

With the assumptions stated above, the wave equations for the electric and magnetic intensities, Eqs. (13.05-4 and 5), reduce to

$$\frac{\partial^2 E_x}{\partial y^2} = \mu\epsilon \frac{\partial^2 E_x}{\partial t^2} \tag{1}$$

$$\frac{\partial^2 H_z}{\partial y^2} = \mu\epsilon \frac{\partial^2 H_z}{\partial t^2} \tag{2}$$

Comparison of these equations with Eqs. (8.01-7 and 8) shows a similarity between the expressions for wave propagation in a lossless dielectric and wave propagation along lossless transmission lines.

A solution of Eqs. (1) and (2) is of the form

$$E_x = Af_1\left(t + \frac{y}{v_c}\right) + Bf_2\left(t - \frac{y}{v_c}\right) \tag{3}$$

The function $f_1[t + (y/v_c)]$ represents a wave traveling in the $-y$ direction, whereas $f_2[t - (y/v_c)]$ represents a wave traveling in the $+y$ direction, both waves having a velocity v_c. The functions f_1 and f_2 are determined by the waveform of the signal radiated from the source.

In a homogeneous medium, there can be an outgoing wave but no reflected wave. Either one of the terms in Eq. (3) may be used to represent the outgoing wave. In order to be consistent with the following discussion, we shall assume that the outgoing wave is traveling in the $-y$ direction. For a sinusoidal time variation, the outgoing wave of electric intensity may be represented by

$$E_x = E'e^{j\omega[t + (y/v_c)]} \tag{4}$$

The magnetic intensity is obtained by inserting Eq. (4) into the curl equation (13.06-1). For the assumed conditions, the curl equation reduces to

$$\frac{\partial E_x}{\partial y} = j\omega\mu H_z \tag{5}$$

Substitution of Eq. (4) gives

$$H_z = \sqrt{\frac{\epsilon}{\mu}}\, E'e^{j\omega[t + (y/v_c)]} = \sqrt{\frac{\epsilon}{\mu}}\, E_x \tag{6}$$

To obtain the velocity of the wave, we may substitute Eq. (4) into (1) above, yielding

$$v_c = \frac{1}{\sqrt{\mu\epsilon}} \tag{7}$$

For a lossless dielectric medium, the velocity is equal to the velocity of light in the given medium. In free space, the velocity is $v_c = 3 \times 10^8$ meters per second.
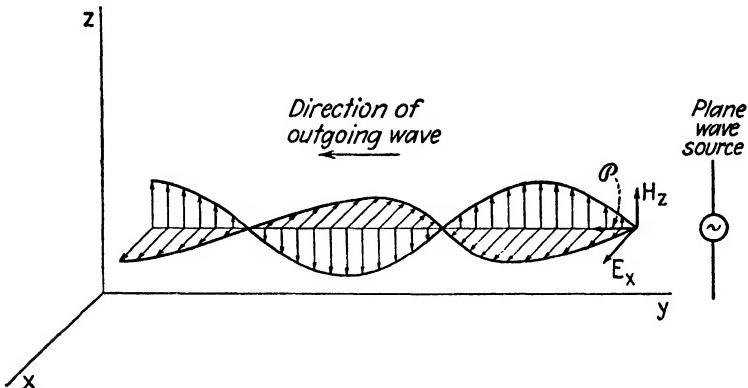
FIG. 1.—Plane wave traveling in the $-y$ direction.

The ratio of electric intensity to magnetic intensity for an outgoing wave is defined as the *intrinsic impedance* of the medium and is designated by the symbol η. Dividing Eq. (4) by (6), we have, for a lossless dielectric medium,

$$\eta = \frac{E_x}{H_z} = \sqrt{\frac{\mu}{\epsilon}} \tag{8}$$

The intrinsic impedance of a medium is analogous to the characteristic impedance of a transmission line. Later we shall derive a more general expression for the intrinsic impedance.

Equations (4) and (6) represent the electric and magnetic intensities of a wave which travels in the $-y$ direction with a velocity equal to the velocity of light. The electric and magnetic intensities are in time phase but in space quadrature. The instantaneous values of E_x and H_z may be evaluated by taking the real part of Eqs. (4) and (6) as explained in the footnote of Sec. 8.02. If E' is real, the real parts of these equations are $E_x = E' \cos \omega[t + (y/v_c)]$ and $H_z = (E'/\eta) \cos \omega[t + (y/v_c)]$. These equations may be used to plot the electric and magnetic intensities as functions of either distance or time. Figure 1 shows the intensities as a function of distance, with time held constant. The Poynting vector, representing power density, is perpendicular to $\bar{E}$ and $\bar{H}$ and is therefore in the $-y$ direction.

14.02. Uniform Plane Waves—General Case.—The preceding section dealt with plane-wave propagation in a lossless dielectric medium. Let us now consider the more general case of uniform plane-wave propagation in any homogeneous isotropic medium.

Again we assume an electric intensity in the x direction and a magnetic intensity in the z direction, with the wave traveling in the y direction. A wave in which the electric and magnetic intensities are both perpendicular to the direction of propagation is known as a *transverse electromagnetic wave* (*TEM* wave). The intensities are assumed to have no variation in the xz plane, and a time variation of the form $e^{j\omega t}$ is assumed. The wave equations (13.06-3 and 4) then become

$$\frac{\partial^2 E_x}{\partial y^2} = \gamma^2 E_x \tag{1}$$

$$\frac{\partial^2 H_z}{\partial y^2} = \gamma^2 H_z \tag{2}$$

where γ is the *intrinsic propagation constant* given by

$$\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)} \tag{13.06-5}$$

We may write the solution of Eq. (1) in the form

$$E_x = E'_R e^{\gamma y} + E''_R e^{-\gamma y} \tag{3}$$

To obtain the magnetic intensity, substitute E_x from Eq. (3) into the curl equation (13.06-1), or into the simplified form, Eq. (14.01-5), giving

$$H_z = \frac{\gamma}{j\omega\mu} (E'_R e^{\gamma y} - E''_R e^{-\gamma y}) \tag{4}$$

For convenience, we write Eq. (4) in the form

$$H_z = H'_R e^{\gamma y} + H''_R e^{-\gamma y} \tag{5}$$

where

$$H'_R = \frac{\gamma}{j\omega\mu} E'_R \qquad \text{and} \qquad H''_R = -\frac{\gamma}{j\omega\mu} E''_R$$

By comparing Eqs. (3) and (4) with (8.02-10 and 11), we again observe a similarity between the relationships for plane-wave propagation and those for waves on transmission lines.

In the following discussion, it will be assumed that the outgoing (or incident) wave travels in the $-y$ direction and that the reflected wave travels in the $+y$ direction. This convention will enable us to write the equations for plane-wave propagation in a form identical to that of the transmission-line equations. According to this convention, the first terms in Eqs. (3) and (4) represent outgoing waves, whereas the second terms

represent reflected waves. The intrinsic impedance of a medium was previously defined as the ratio of electric intensity to magnetic intensity for an outgoing wave. Upon applying this definition to Eqs. (3) and (5), we obtain the ratio of electric to magnetic intensity for either the outgoing or the reflected wave

$$\eta = \frac{E'_R}{H'_R} = -\frac{E''_R}{H''_R} = \frac{j\omega\mu}{\gamma} \tag{6}$$

If we had derived the magnetic intensity by inserting Eq. (3) into (13.06-2) instead of (13.06-1), a relationship similar to Eq. (4) would have resulted, but with $\eta = \gamma/(\sigma + j\omega\epsilon)$. Therefore the intrinsic impedance for *TEM* waves may be represented by either of the relationships

$$\eta = \frac{j\omega\mu}{\gamma} = \frac{\gamma}{\sigma + j\omega\epsilon} \tag{7}$$

14.03. Intrinsic Impedance and Propagation Constant.—The intrinsic impedance and intrinsic propagation constant are properties of the medium. They are analogous to the characteristic impedance and propagation constant, respectively, of transmission lines. The expressions for these quantities may be simplified for the special cases where the medium is (1) a lossless dielectric or (2) a good conductor.

First, however, let us rewrite the complete expressions for ready reference

$$\eta = \frac{j\omega\mu}{\gamma} = \frac{\gamma}{\sigma + j\omega\epsilon} = \sqrt{\frac{j\omega\mu}{\sigma + j\omega\epsilon}} \tag{14.02-7}$$

$$\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)} = \alpha + j\beta \tag{13.06-5}$$

Consider now the case of a lossless dielectric medium, for which we have $\sigma = 0$. The intrinsic impedance and propagation constant then reduce to

$$\eta = \sqrt{\frac{\mu}{\epsilon}} \tag{1}$$

$$\gamma = j\omega\sqrt{\mu\epsilon} \tag{2}$$

The intrinsic impedance of a lossless dielectric is of the nature of a pure resistance and has a value of 376.6 ohms for free space. Equation (1) resembles the familiar form of the characteristic impedance of a lossless line, $Z_0 = \sqrt{L/C}$. Since all dielectric mediums have approximately the same permeability ($\mu_0 = 4\pi \times 10^{-7}$) and there are no known dielectrics having permittivities appreciably less than that of free space, it follows that the intrinsic impedance of free space is about the maximum attainable value for known dielectric materials.

The intrinsic propagation constant of lossless dielectrics is imaginary; consequently, we have $\alpha = 0$ and $\beta = \omega\sqrt{\mu\epsilon}$. These relationships also have their counterpart in the lossless transmission line where we find $\alpha = 0$ and $\beta = \omega\sqrt{LC}$. The wavelength and phase velocity for plane waves are

$$\lambda = \frac{2\pi}{\beta} = \frac{2\pi}{\omega\sqrt{\mu\epsilon}} \tag{3}$$

$$v_c = f\lambda = \frac{1}{\sqrt{\mu\epsilon}} \tag{4}$$

Now consider the special case in which the medium is a good conductor. In a good conductor we find that $\sigma \gg \omega\epsilon$. This is valid over the entire range of frequencies extending from the audio frequencies through the microwave frequencies. As an example, consider silver, having a conductivity of $\sigma = 6.14 \times 10^7$ mhos per meter. While the permittivity of metals is not accurately known, the evidence indicates that it is of the same order of magnitude as the permittivity of free space. Assuming that $\epsilon = \epsilon_0$, we find that even at the relatively high frequency of 10^{11} cycles per second, the value of $\omega\epsilon$ is of the order of $\omega\epsilon = 5$ as compared with $\sigma = 6.14 \times 10^7$. It is therefore apparent that we can assume that $\sigma \gg \omega\epsilon$ for good conductors. Equations (13.06-5) and (14.02-7) then reduce to

$$\gamma = \sqrt{j\omega\mu\sigma} = \sqrt{\frac{\omega\mu\sigma}{2}} + j\sqrt{\frac{\omega\mu\sigma}{2}} \tag{5}$$

$$\eta = \sqrt{\frac{\omega\mu}{2\sigma}}(1 + j) = \sqrt{\frac{\omega\mu}{\sigma}} \angle 45° \tag{6}$$

The attenuation constant, phase constant, wavelength, and phase velocity may be obtained from Eq. (5) as follows:

$$\alpha = \beta = \sqrt{\frac{\omega\mu\sigma}{2}} \tag{7}$$

$$\lambda = \frac{2\pi}{\beta} = 2\pi\sqrt{\frac{2}{\omega\mu\sigma}} \tag{8}$$

$$v = f\lambda = \sqrt{\frac{2\omega}{\mu\sigma}} \tag{9}$$

The intrinsic impedance of conductors is extremely small in comparison with that of most dielectric mediums. This indicates a low ratio of electric to magnetic intensity in conductors. Equation (6) shows that the electric intensity leads the magnetic intensity by a time phase angle of 45 degrees

in conductors. The attenuation and phase constants have equal values in good conductors.

Some idea of the properties of a wave in metal may be obtained by evaluating the intrinsic impedance, wavelength, phase velocity, and attenuation constant in silver at a frequency of 100 megacycles. At this frequency, the intrinsic impedance of free space is $\eta = 376.6$ ohms, whereas that of silver is $\eta = 0.0025 + j0.0025$ ohm. The wavelength and phase velocity in free space are $\lambda = 3$ meters and $v_c = 3 \times 10^8$ meters per second, respectively. In contrast with these values, we obtain for silver $\lambda = 0.004$ centimeter and $v = 4 \times 10^3$ meters per second. Thus, the wavelength and phase velocity in a conductor are very much smaller than the corresponding values in free space. The attenuation constant is $\alpha = 15.7 \times 10^4$ nepers per meter or 136×10^4 decibels per meter, which represents an extremely high attenuation. The wave is therefore attenuated to a negligible value in a distance of a few thousandths of a centimeter.

14.04. Power Flow.—Poynting's vector, $\overline{\mathcal{P}} = \overline{E} \times \overline{H}$, may be used to evaluate the instantaneous value of power density in an electromagnetic wave. In the plane-wave example of the preceding article, the intensity vectors are mutually perpendicular and the scalar value of Poynting's vector may therefore be written $\mathcal{P} = EH$.

Since we will be dealing largely with fields having sinusoidal time variation, it will be convenient to have an expression for the time-average power density. While we could derive such a relationship on rigorous grounds, the same results may be more readily obtained by analogy with transmission-line equations. The time-average power flow at any point on a transmission line is $P_{\text{av}} = \frac{1}{2} VI \cos \theta$, where V and I are peak values and θ is the time phase angle between V and I. By analogy, the time-average power density in a uniform plane wave is $\mathcal{P}_{\text{av}} = \frac{1}{2} |E| |H| \cos \theta$, where $|E|$ and $|H|$ are the peak values of the electric and magnetic intensities, and θ is the time phase angle between E and H. This may also be shown to be equivalent to

$$\mathcal{P}_{\text{av}} = \frac{1}{2} \operatorname{Re} (EH^*) \tag{1}$$

where E and H are complex values, and H^* is the complex conjugate of H. The complex conjugate is formed by reversing the sign of the phase angle. Thus, if $A = |A_0| e^{j\theta}$, we have $A^* = |A_0| e^{-j\theta}$. The symbol Re in Eq. (1) signifies that the real part of the bracketed term is to be retained and the imaginary part is to be discarded.

For an outgoing wave only, we have $E/H = \eta$. Replacing E in Eq. (1) by ηH, we obtain

$$\mathcal{P}_{\text{av}} = \frac{1}{2} \operatorname{Re} (HH^*\eta) = \frac{|H|^2}{2} \operatorname{Re} (\eta) \tag{2}$$

The later form follows from the fact that $HH^* = |H|^2$.

For a lossless dielectric medium, we have $\eta = \sqrt{\mu/\epsilon}$ and Eq. (2) reduces to

$$\mathcal{P}_{\mathrm{av}} = \frac{|H|^2 \eta}{2} = \frac{|E|^2}{2\eta} \tag{3}$$

For a good conductor, we have $\eta = (1 + j)\sqrt{\omega\mu/2\sigma}$ and Eq. (3) becomes

$$\mathcal{P}_{\mathrm{av}} = \frac{|H|^2}{2} \sqrt{\frac{\omega\mu}{2\sigma}} \tag{4}$$

Equations (2), (3), and (4) are the most convenient forms to use for evaluating the time-average power density. In these equations $|E|$ and $|H|$ are peak values of the intensities.

14.05. Plane-wave Reflection at Normal Incidence.[1,2]—When an electromagnetic wave, traveling in one medium, impinges upon a boundary-surface between two mediums having different intrinsic impedances, a partial reflection occurs at the boundary between the mediums. This results in a reflected wave traveling back toward the source in the first medium, and a transmitted wave in the second medium. In the first medium the incident and reflected waves combine to produce a standing wave. If the two mediums have approximately the same intrinsic impedances, most of the wave energy is transmitted into the second medium and the reflected wave is relatively small. Conversely, if the intrinsic impedances differ greatly, the transmitted wave is small, and the reflected wave is relatively large. The standing-wave ratio in the first medium (ratio of maximum to minimum standing wave of electric intensity) may be used as a measure of the degree of impedance mismatch.

Consider a uniform plane wave which is normally incident upon a plane surface between two mediums, designated by the subscripts 1 and 2 in Fig. 2. Both mediums are assumed to be infinite in extent in all directions except at the boundary surface. This surface is chosen to coincide with the plane $y = 0$ in Fig. 2. The mediums are assumed to be homogeneous but they may be conducting, semiconducting, or insulating. The incident wave travels in the $-y$ direction and contains the propagation term $e^{\gamma y}$ while the reflected wave travels in the $+y$ direction with a propagation term $e^{-\gamma y}$. The transmission-line analogue consists of a line having parameters Z_{01} and γ_1 terminated by a second line which is infinitely long and which has the parameters Z_{02} and γ_2.

[1] Schelkunoff, S. A., The Impedance Concept and Its Application to Problems of Reflection, Refraction, Shielding, and Power Absorption, *Bell System Tech. J.*, vol. 17, pp. 17–48; January, 1938.

[2] Schelkunoff, S. A., "Electromagnetic Waves," D. Van Nostrand Company, Inc., New York, 1943.

The electric and magnetic intensities in medium 1 are given by Eqs. (14.02-3 and 4). With the substitution of Eq. (14.02-6), these may be written

$$E_x = E'_R e^{\gamma_1 y} + E''_R e^{-\gamma_1 y} \qquad (14.02\text{-}3)$$

$$H_z = \frac{E'_R}{\eta_1} e^{\gamma_1 y} - \frac{E''_R}{\eta_1} e^{-\gamma_1 y} \qquad (14.02\text{-}4)$$

where E'_R and E''_R are the incident and reflected wave intensities, re spectively, at the surface $y = 0$.
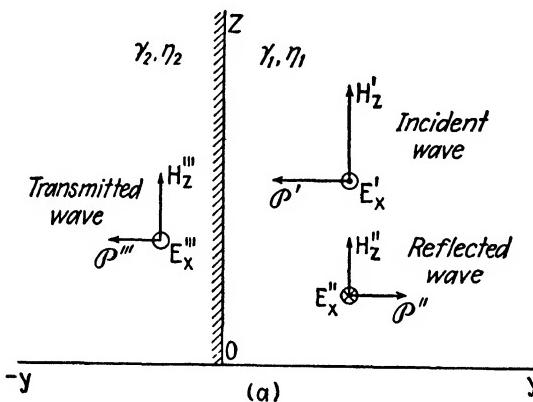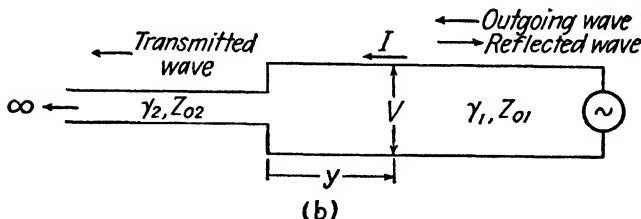

FIG. 2.—Intensities of a normally incident plane wave and the transmission-line analogy.

Referring to the transmission-line in Fig. 2b, the impedance terminating line 1 is the characteristic impedance of line 2, or Z_{02}. Likewise, in Fig. 2a, the wave impedance which terminates medium 1 is the intrinsic impedance of medium 2, or η_2. The same conclusion can be reached by applying the boundary conditions of Sec. 13.08. These require that the tangential components of E and H be equal on either side of the boundary. Consequently, the ratio E/H must be the same on either side of the boundary. The wave in medium 2 is an outgoing wave only; hence $E'''_x/H'''_z = \eta_2$. Therefore, to satisfy the boundary conditions, we must have, at the boundary in medium 1, $E_R/H_R = E'''_x/H'''_z = \eta_2$, where E_R and H_R are the resultant

electric and magnetic intensities at the surface in medium 1 (the sum of the incident and reflected-wave intensities).   Thus, the impedance terminating medium 1 is the intrinsic impedance of medium 2.

At the boundary surface in medium 1 $y = 0$, Eqs. (14.02-3 and 4) become $E_R = E_R' + E_R''$ and $H_R = (1/\eta_1)(E_R' - E_R'')$. Substituting $H_R = E_R/\eta_2$ into the second of these equations and solving them for E_R' and E_R'' we obtain

$$E_R' = \frac{E_R}{2} \left(1 + \frac{\eta_1}{\eta_2} \right) \qquad E_R'' = \frac{E_R}{2} \left(1 - \frac{\eta_1}{\eta_2} \right) \tag{1}$$

Upon inserting these into (14.02-3 and 4), we obtain the intensity equations

$$E_x = \frac{E_R}{2} \left(1 + \frac{\eta_1}{\eta_2} \right) e^{\gamma_1 y} + \frac{E_R}{2} \left(1 - \frac{\eta_1}{\eta_2} \right) e^{-\gamma_1 y} \tag{2}$$

$$H_z = \frac{E_R}{2\eta_1} \left(1 + \frac{\eta_1}{\eta_2} \right) e^{\gamma_1 y} - \frac{E_R}{2\eta_1} \left(1 - \frac{\eta_1}{\eta_2} \right) e^{-\gamma_1 y} \tag{3}$$

These equations are similar to the transmission-line equations (8.02-16 and 17). They may be expressed in hyperbolic form similar to Eqs. (8.02-21 to 23), as follows:

$$E_x = E_R \left(\cosh \gamma_1 y + \frac{\eta_1}{\eta_2} \sinh \gamma_1 y \right) \tag{4}$$

$$H_z = \frac{E_R}{\eta_2} \left(\cosh \gamma_1 y + \frac{\eta_2}{\eta_1} \sinh \gamma_1 y \right) \tag{5}$$

We now define a *wave impedance Z* as the ratio of the *resultant* electric intensity to the *resultant* magnetic intensity, at any given point. This is analogous to the impedance at any point on a transmission line,

$$Z = \frac{E_x}{H_z} = \eta_1 \left(\frac{\eta_2 + \eta_1 \tanh \gamma_1 y}{\eta_1 + \eta_2 \tanh \gamma_1 y} \right) \tag{6}$$

In writing the equations for the intensities in medium 2, we recall that the tangential electric intensities are equal at the boundary. Since E_R is the tangential intensity in medium 1, it must also be the surface intensity in medium 2. Therefore the intensities in medium 2 are

$$E_x''' = E_R e^{\gamma_2 y} \tag{7}$$

$$H_z''' = \frac{E_R}{\eta_2} e^{\gamma_2 y} \tag{8}$$

The *reflection coefficient* [1] r_R is defined as the ratio of the electric intensity of the reflected wave to that of the incident wave at the boundary surface, or $r_R = E_R''/E_R'$.   Inserting the values of E_R'' and E_R' from Eq. (1) yields

$$r_R = \frac{E_R''}{E_R'} = \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1} \tag{9}$$

It is also convenient to define a *transmission coefficient* as the ratio of electric intensity in medium 2 to the electric intensity of the incident wave in medium 1, both being taken at the reflecting surface, or $r_T = E_R/E_R'$. Replacing E_R' by Eq. (1), we obtain

$$r_T = \frac{E_R}{E_R'} = \frac{2\eta_2}{\eta_2 + \eta_1} \tag{10}$$

The transmission coefficient enables us to evaluate the electric intensity in medium 2 in terms of the electric intensity of the outgoing wave in medium 1.

The expressions in reflection coefficient form, analogous to Eqs. (2) to (6), are

$$E_x = E_R' e^{\gamma_1 y}(1 + r_R \epsilon^{-2\gamma_1 y}) \tag{11}$$

$$H_z = \frac{E_R'}{\eta_1} e^{\gamma_1 y}(1 - r_R \epsilon^{-2\gamma_1 y}) \tag{12}$$

$$Z = \frac{E_x}{H_z} = \eta_1 \left(\frac{1 + r_R \epsilon^{-2\gamma_1 y}}{1 - r_R \epsilon^{-2\gamma_1 y}} \right) \tag{13}$$

14.06. Normal-incidence Reflection from a Conductor.—As a special case of the foregoing relationships, let us assume that medium 2 is a perfect conductor.   The intrinsic impedance of a perfect conductor is zero and there can be no electric or magnetic fields inside the conductor.   The reflection coefficient is $r_R = -1$ and the transmission coefficient is $r_T = 0$, indicating total reflection of the incident wave.   The boundary conditions require that the electric intensity E_R be zero.   However, the magnetic intensity H_R at the boundary in medium 1 is not zero.   In Eqs. (14.05-4, 5, and 6) we substitute $E_R = 0$ and $\eta_2 = 0$.   To eliminate the indeterminant, we use $H_R = E_R/\eta_2$, yielding

$$E_x = H_R \eta_1 \sinh \gamma_1 y \tag{1}$$

$$H_z = H_R \cosh \gamma_1 y \tag{2}$$

$$Z = \frac{E_x}{H_z} = \eta_1 \tanh \gamma_1 y \tag{3}$$

---

[1] In optics the reflection coefficient is taken as the ratio $r_R = H_R''/H_R'$.   This reflection coefficient is equal in magnitude to that of Eq. (9) but has opposite sign.   The definition given by Eq. (9) is consistent with the definition of the reflection coefficient for the transmission line.

These equations are similar to those of the short-circuited line discussed in Sec. 8.07. The curves of $\left|\, V/I_R Z_0 \,\right|$ and $\left|\, I/I_R \,\right|$ in Fig. 7 may be used to represent $\left|\, E_x/H_R\eta_1 \,\right|$ and $\left|\, H_z/H_R \,\right|$, respectively. The impedance curves of Fig. 8, Chap. 8, also represent the wave impedance ratio Z/η_1, as given by Eq. (3).

If medium 1 is lossless and medium 2 is a perfect conductor, we have $\gamma_1 = j\beta_1$, and the hyperbolic functions in Eqs. (1) to (3) reduce to trigonometric functions, yielding

$$E_x = jH_R\eta_1 \sin \beta_1 y \tag{4}$$

$$H_z = H_R \cos \beta_1 y \tag{5}$$

$$Z = j\eta_1 \tan \beta_1 y \tag{6}$$

The incident and reflected waves combine in such a manner as to produce the standing waves shown in Fig. 3. The electric and magnetic fields have
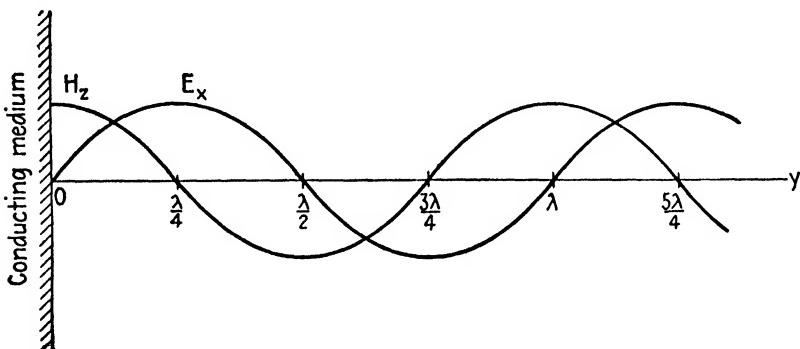

FIG. 3.—Standing waves produced by reflection from a perfect conductor.

a sinusoidal time variation and the values plotted in Fig. 3 represent the peak values of the sine wave. The electric and magnetic intensities are in time quadrature as well as in space quadrature.

At the conducting surface E_R is zero, while H_R is a maximum, having a value of twice the magnetic intensity of the incident wave. The magnetic field is terminated by a current flowing on the geometrical surface of the conductor (for a perfect conductor). This current flows in a direction perpendicular to the magnetic intensity. The value of the surface current density, as given by Eq. (13.08-5), is $J_s = H_R$.

The standing wave of electric intensity has its maximum values at distances $y = n\lambda/4$ from the reflecting surface, where n is an odd integer. Nodal values of electric intensity occur at distances of $y = n\lambda/4$, where n is an even integer.

14.07. Depth of Penetration and Skin-effect Resistance.—If the conducting medium has finite conductivity, a very small portion of the energy

of the incident wave enters the conductor. The remaining energy is reflected at the surface of the conductor. The wave entering the conductor is rapidly attenuated as it travels inward from the surface, the intensities being attenuated by the factor [1] $e^{-\alpha_2 y}$. Consequently, the electric and magnetic intensities decrease to a value of $1/e$, or 36.8 per cent of the surface value at a depth for which $\alpha_2 y = 1$. This particular value of y is known as the *depth of penetration* or *skin depth*, and is represented by the symbol δ_2. From Eq. (14.03-7) we obtain

$$\delta_2 = \frac{1}{\alpha_2} = \sqrt{\frac{2}{\omega \mu_2 \sigma_2}} \qquad (1)$$

In silver at a frequency of 100 megacycles, the depth of penetration is $\delta_2 = 0.637 \times 10^{-5}$ meter. Thus, in good conductors at microwave frequencies, the wave is attenuated to a negligible value within a few thousandths of a centimeter.

The transmission coefficient enables us to evaluate the intensities in the conductor in terms of the electric intensity of the incident wave. Since we have $\eta_1 \gg \eta_2$, the transmission coefficient, from Eq. (14.05-10), becomes $r_T = 2\eta_2/\eta_1$, and the electric intensity at the surface in either medium is

$$E_R = \frac{2\eta_2}{\eta_1} E_R' \qquad (2)$$

and the magnetic intensity is

$$H_R = \frac{E_R}{\eta_2} = \frac{2E_R'}{\eta_1}$$

The time-average power density entering the surface of the conductor is obtained from Eq. (14.04-4), thus

$$\mathcal{P}_{\text{av}} = \frac{|H_R|^2}{2} \sqrt{\frac{\omega \mu_2}{2\sigma_2}} \qquad (3)$$

In an a-c circuit, the resistance may be defined by the relationship $P_{\text{av}} = I^2 R/2$, or $R = 2P_{\text{av}}/I^2$, where P_{av} is the time-average power consumed in the resistance and I is the peak value of current flowing through the resistance. The skin-effect resistance for the plane conducting surface may be defined in a similar manner by the relationship $\mathcal{R}_s = 2\mathcal{P}_{\text{av}}/I^2$, where $\mathcal{P}_{\text{av}}$ is the power density as given by Eq. (3). The current I is the peak value of the current flowing through a cross section of the conductor taken parallel to the yz plane in Fig. 2, with unit length in the z direction

---

[1] A wave traveling in the $+y$ direction will contain the attenuation term $e^{-\alpha y}$. A wave traveling in the $-y$ direction has an attenuation term $e^{\alpha y}$, but the sign of y reverses. Consequently, the intensities may be considered to be attenuated by an amount $e^{-\alpha y}$ for either direction of travel.

and infinite thickness in the y direction. This current is equal to the surface value of the magnetic intensity, or $I = H_R$. Inserting this value of current, together with $\mathcal{P}_{av}$ from Eq. (3), into the equation for $\mathcal{R}_s$ above, we obtain

$$\mathcal{R}_s = \sqrt{\frac{\omega \mu_2}{2\sigma_2}} = \frac{1}{\delta_2 \sigma_2} \tag{4}$$

where δ_2 is the depth of penetration. The quantity $\delta_2 \sigma_2$ may be viewed as the conductance of a slab of conductor of unit length, unit width, and depth equal to δ_2. The skin-effect resistance is the reciprocal of this conductance. It is also interesting to observe that the intrinsic impedance as given by Eq. (14.03-6) may be written in the form $\eta_2 = (1 + j)\mathcal{R}_s$. The real part of the intrinsic impedance is the skin-effect resistance.

14.08. Normal-incidence Reflection from a Lossless Dielectric.—If both mediums are lossless dielectrics having different permittivities, partial reflection occurs at the boundary surface. The reflection and transmission coefficients are obtained by substituting $\eta_2 = \sqrt{\mu_2/\epsilon_2}$ and $\eta_1 = \sqrt{\mu_1/\epsilon_1}$ into Eqs. (14.05-9 and 10). Since dielectric materials are nonmagnetic, their permeabilities are equal. The reflection and transmission coefficients then become

$$r_R = \frac{E_R''}{E_R'} = \frac{\sqrt{(\epsilon_1/\epsilon_2)} - 1}{\sqrt{(\epsilon_1/\epsilon_2)} + 1} \tag{1}$$

$$r_T = \frac{E_R}{E_R'} = \frac{2}{1 + \sqrt{\epsilon_2/\epsilon_1}} \tag{2}$$

The reflection coefficient is zero when $\epsilon_1/\epsilon_2 = 1$, that is, when the two dielectric mediums have the same permittivities. As the ratio departs farther from unity value, the reflection coefficient increases and the transmission coefficient decreases, indicating increasing reflection at the boundary surface.

14.09. Multiple Reflection and Impedance Matching.—Problems dealing with multiple reflection at normal incidence may be analyzed by the foregoing methods. As an example, consider the arrangement shown in Fig. 4, consisting of three different mediums, represented by subscripts 1, 2, and 3. Let us evaluate the wave impedance at the surface between mediums 1 and 2. In the transmission-line analogy, this impedance corresponds to the impedance looking to the left at points ab, in Fig. 4b. At this point we have the impedance looking into transmission line 2 terminated by an infinitely long line 3. Equation (14.05-6) may be modified to express the equivalent wave impedance in Fig. 4a, thus

$$Z = \frac{E_x}{H_z} = \eta_2 \left(\frac{\eta_3 + \eta_2 \tanh \gamma_2 l_2}{\eta_2 + \eta_3 \tanh \gamma_2 l_2} \right) \tag{1}$$

This is the wave impedance terminating medium 1. The reflection and transmission coefficients at the surface between mediums 1 and 2 are obtained by substituting the value of Z from Eq. (1) into Eqs. (14.05-9 and 10), thus

$$r_R = \frac{Z - \eta_1}{Z + \eta_1} \tag{2}$$

$$r_T = \frac{2Z}{Z + \eta_1} \tag{3}$$

An interesting application of multiple reflection is that of matching the wave impedances of two different dielectric mediums. In our transmission-line theory we found that unequal generator and load impedances could be
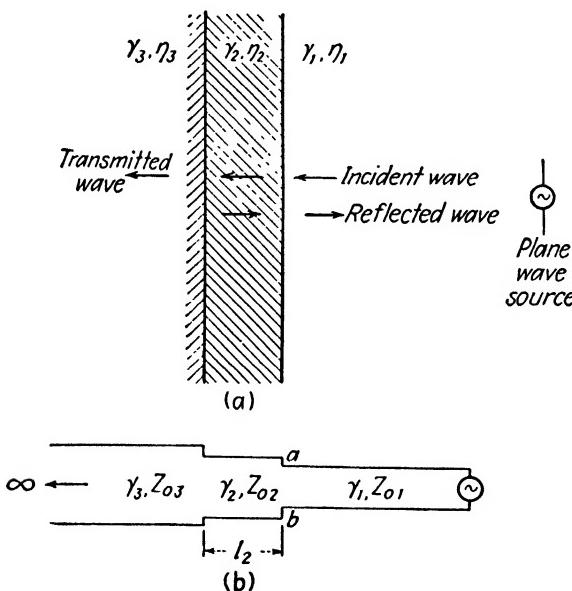

FIG. 4.—Multiple reflection and the transmission-line analogy.

matched, if they are both resistances, by inserting a quarter-wavelength section of line, having the proper value of characteristic impedance, between the generator and load. This results in an impedance match at the generator and maximum power transfer from the generator to the load. In a similar manner, a slab of dielectric, which is a quarter-wavelength thick, with properly chosen intrinsic impedance, may be inserted between two different dielectric mediums to obtain a match of the intrinsic impedances. For an impedance match, there will be no standing waves in medium 1 and maximum power will be transferred from medium 1 to medium 3, despite

the fact that the two mediums have different intrinsic impedances. The absence of standing waves in medium 1 may be attributed to the fact that there are two reflected waves in this medium which are equal in amplitude but 180 degrees out of phase, thereby canceling each other. The two reflected waves originate at the surfaces between mediums 1 and 2 and between mediums 2 and 3.

Another way of viewing this is that medium 2 serves as an impedance transformer, which is chosen so that medium 1 is terminated in a wave impedance equal to its intrinsic impedance. The impedance terminating medium 1 is expressed by Eq. (1). Since medium 2 is a lossless dielectric, we may replace $\tanh \gamma_2 l_2$ by $j \tan \beta_2 l_2$. If medium 2 is a quarter-wavelength thick, we have $\beta_2 l_2 = \pi/2$ and Eq. (1) reduces to $Z = \eta_2^2/\eta_3$. For an impedance match, we must have $Z = \eta_1$, or

$$\eta_1 = \frac{\eta_2^2}{\eta_3} \tag{4}$$

This is similar to Eq. (10.09-2) for the quarter-wavelength line used as an impedance transformer. Replacing η_1, η_2, and η_3 by their respective values for a dielectric medium, we obtain

$$\frac{\mu_2}{\epsilon_2} = \sqrt{\frac{\mu_1 \mu_3}{\epsilon_1 \epsilon_3}} \tag{5}$$

Since the permeabilities are all equal, this may be written

$$\epsilon_2 = \sqrt{\epsilon_1 \epsilon_3} \tag{6}$$

Hence, the permittivity of medium 2, required for maximum power transfer, is the geometrical mean between the permittivities of mediums 1 and 3.

14.10. Oblique-incidence Reflection—Polarization Normal to the Plane of Incidence.—The foregoing discussion has dealt with the reflection of uniform-plane waves impinging upon a plane boundary surface at normal incidence. Let us now consider plane-wave reflection at oblique incidence.

Consider the reference axis shown in Fig. 5. The boundary surface between the two mediums is assumed to coincide with the plane $y = 0$. Poynting's vectors for the incident, reflected, and refracted waves are assumed to lie in the yz plane, this plane being referred to as the *plane of incidence*. Two special cases are considered: (1) the electric intensity normal to the plane of incidence, as shown in Fig. 5, and (2) the electric intensity parallel to the plane of incidence. These will be referred to as waves polarized *normal* to the plane of incidence and polarized *parallel* to the plane of incidence, respectively.[1] Later, in the treatment of wave

---

[1] The direction of polarization is taken as the direction of the electric intensity vector.

guides, the first case will be identified as a transverse-electric (*TE*) wave and the second as a transverse-magnetic (*TM*) wave.

Consider first a uniform plane wave with polarization normal to the plane of incidence. The general propagation term for a wave traveling with respect to a rectangular coordinate system is $e^{\pm\gamma(lx+my+nz)}$, where $l = \cos\theta_x$, $m = \cos\theta_y$, and $n = \cos\theta_z$ are the direction cosines. The angles θ_x, θ_y, and θ_z are the angles between the wave normal and the x, y, and z axes, respectively.
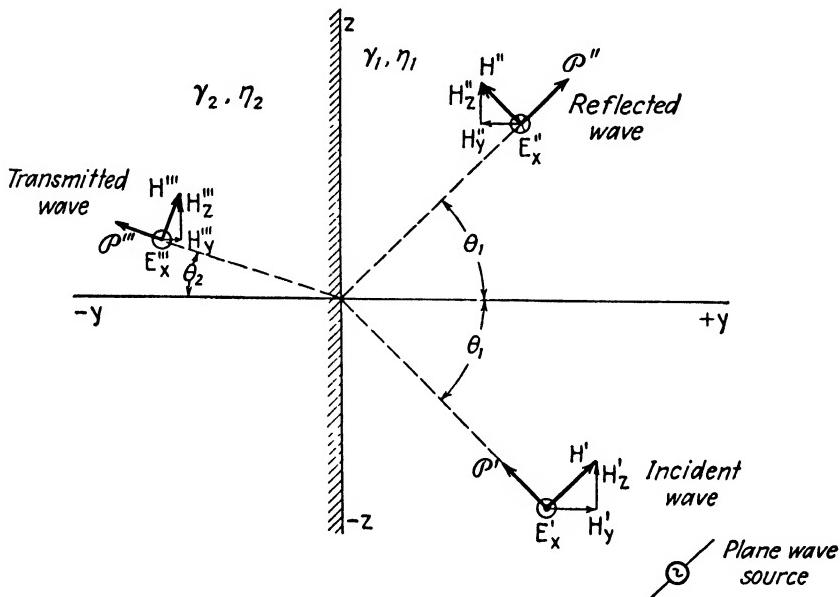

FIG. 5.—Intensity components of a plane wave polarized normal to the plane of incidence.

For the incident wave of Fig. 5, we have $l = 0$, $m = \cos\theta_1$, $n = -\sin\theta_1$. Since the wave is traveling toward the origin, the positive sign is used in the propagation term. The incident wave is therefore

$$E'_x = E'_R e^{\gamma_1(y\cos\theta_1 - z\sin\theta_1)} \tag{1}$$

where E'_R is the electric intensity of the incident wave at the origin.

For the reflected wave, we have $l = 0$, $m = \cos\theta_1$, $n = \sin\theta_1$. This wave is traveling away from the origin; hence we use a negative sign, yielding

$$E''_x = E''_R e^{-\gamma_1(y\cos\theta_1 + z\sin\theta_1)} \tag{2}$$

where E''_R is the intensity of the reflected wave at the origin. The magnetic-intensity components may be obtained by substituting the expression for E_x into the curl equation (13.06-1), or they may be obtained by dividing

the incident wave electric intensity by η_1 and the reflected wave by $-\eta_1$. The resultant electric and magnetic intensities in medium 1, therefore, are

$$E_x = (E_R' e^{\gamma_1 y \cos \theta_1} + E_R'' e^{-\gamma_1 y \cos \theta_1}) e^{-\gamma_1 z \sin \theta_1} \tag{3}$$

$$H' = \frac{E_R'}{\eta_1} e^{\gamma_1 y \cos \theta_1} e^{-\gamma_1 z \sin \theta_1}$$

$$H'' = -\frac{E_R''}{\eta_1} e^{-\gamma_1 y \cos \theta_1} e^{-\gamma_1 z \sin \theta_1} \tag{4}$$

A uniform plane wave at oblique incidence may be interpreted in several ways. We may view the wavefront as an equiphase plane having uniform electric and magnetic intensities over its surface. This wave impinges upon the boundary surface at oblique incidence, causing a plane reflected wave in medium 1 and a plane refracted wave in medium 2. Another equally valid interpretation is to view the wave as consisting of a standing wave in a direction *normal* to the boundary surface and a traveling wave moving in a direction *parallel* to the boundary surface.

The field intensity components contributing to the normal components of Poynting's vector are E_x and H_z. To obtain H_z, we multiply Eq. (4) by $\cos \theta_1$, thus,

$$H_z = \frac{\cos \theta_1}{\eta_1} (E_R' e^{\gamma_1 y \cos \theta_1} - E_R'' e^{-\gamma_1 y \cos \theta_1}) e^{-\gamma_1 z \sin \theta_1} \tag{5}$$

A similar expression is obtained for H_y by multiplying Eq. (4) by $\sin \theta_1$.

A few simplifications will now make the equations for E_x and H_z resemble more closely the preceding equations for normal incidence. We define the following quantities

$$Z_{n1} = \eta_1 \sec \theta_1 \tag{6}$$

$$\gamma_{n1} = \gamma_1 \cos \theta_1 \tag{7}$$

$$Z_R = \frac{E_R}{H_{zR}} \tag{8}$$

The foregoing equations express the intensities resulting from reflection at oblique incidence. In these equations, Z_{n1} is the *characteristic wave impedance*, analogous to the intrinsic impedance, and γ_{n1} is the propagation constant, both of these being taken in a direction normal to the reflecting surface. The wave impedance Z_R in Eq. (8) is the ratio of the tangential components of electric and magnetic intensity at the surface, and is therefore the terminating impedance in a direction normal to the reflecting surface.

The values of incident and reflected intensities E_R' and E_R'' at the origin are obtained in terms of E_R by setting $y = z = 0$, with $E_x = E_R$ and $H_z = H_{zR}$ in Eqs. (3) and (5). The additional substitution of Eq. (8) yields $E_R' = E_R/2[1 + (Z_{n1}/Z_R)]$ and $E_R'' = E_R/2[1 - (Z_{n1}/Z_R)]$. Replacing these in the equations for E_x and H_z, we obtain

$$E_x = \left[\frac{E_R}{2} \left(1 + \frac{Z_{n1}}{Z_R} \right) e^{\gamma_{n1} y} + \frac{E_R}{2} \left(1 - \frac{Z_{n1}}{Z_R} \right) e^{-\gamma_{n1} y} \right] e^{-\gamma_1 z \sin \theta_1} \qquad (9)$$

$$H_z = \left[\frac{E_R}{2Z_{n1}} \left(1 + \frac{Z_{n1}}{Z_R} \right) e^{\gamma_{n1} y} - \frac{E_R}{2Z_{n1}} \left(1 - \frac{Z_{n1}}{Z_R} \right) e^{-\gamma_{n1} y} \right] e^{-\gamma_1 z \sin \theta_1} \qquad (10)$$

Comparing these with Eqs. (14.05-2 and 3), we find that the bracketed term in Eqs. (9) and (10) may be interpreted as a wave which is normally incident upon the boundary surface, with a propagation constant γ_{n1} and characteristic wave impedance Z_{n1}.

The incident and reflected waves combine to produce a standing wave in a direction normal to the reflecting surface in medium 1. The term $e^{-\gamma_1 z \sin \theta_1}$ may be viewed as a phase-amplitude variation of the intensities in the z direction.

Consider now the wave in medium 2. Boundary conditions require equality of tangential components of electric intensity on either side of the boundary. Since the tangential electric intensity in medium 1 is E_R, this is also the tangential intensity in medium 2. In medium 2, we have $l = 0$, $m = -\cos \theta_2$, and $n = \sin \theta_2$, and the wave travels away from the origin. Therefore, the intensities are

$$E_x''' = E_R e^{-\gamma_2 (-y \cos \theta_2 + z \sin \theta_2)} \qquad (11)$$

$$H_z''' = \frac{E_R \cos \theta_2}{\eta_2} e^{-\gamma_2 (-y \cos \theta_2 + z \sin \theta_2)} \qquad (12)$$

We let

$$Z_{n2} = \eta_2 \sec \theta_2 \qquad (13)$$

$$\gamma_{n2} = \gamma_2 \cos \theta_2 \qquad (14)$$

yielding,

$$E_x''' = E_R e^{\gamma_{n2} y} e^{-\gamma_2 z \sin \theta_2} \qquad (15)$$

$$H_z''' = \frac{E_R}{Z_{n2}} e^{\gamma_{n2} y} e^{-\gamma_2 z \sin \theta_2} \qquad (16)$$

It still remains to evaluate the wave impedance Z_R which terminates medium 1. Substituting $y = z = 0$ in Eqs. (15) and (16) and using Eq. (8), we obtain $E_x'''/H_z''' = E_R/H_{zR} = Z_{n2}$. Hence, the impedance terminating medium 1 is the characteristic wave impedance of the normally incident components in medium 2. We may therefore substitute Z_{n2} for Z_R in Eqs. (9) and (10). The reflection and transmission coefficients, similar to Eqs. (14.05-9 and 10), are

$$r_R = \frac{E_R''}{E_R'} = \frac{Z_{n2} - Z_{n1}}{Z_{n2} + Z_{n1}} \qquad (17)$$

$$r_T = \frac{E_R}{E_R'} = \frac{2Z_{n2}}{Z_{n2} + Z_{n1}} \qquad (18)$$

If medium (2) is a good conductor, then $Z_{n2} \cong 0$; hence, $r_R \cong -1$, and the angle θ_2 can be assumed to be zero degrees.

14.11. Oblique-incidence Reflection—Polarization Parallel to the Plane of Incidence.—The treatment of the case of polarization parallel to the plane of incidence is similar to that of polarization normal to the plane of incidence except that the electric and magnetic intensities are interchanged. Thus, if the magnetic intensity is H_x, the electric intensity components are E_y and E_z. The intensity components contributing to the normal component of the Poynting vector are E_z and H_x. The intensity equations in medium 1, expressed in terms of the incident and reflected wave magnetic intensities at the origin H'_R and H''_R are

$$H_x = (H'_R e^{\gamma_1 y \cos \theta_1} + H''_R e^{-\gamma_1 y \cos \theta_1}) e^{-\gamma_1 z \sin \theta_1} \qquad (1)$$

Inserting H_x into the curl equation (13.06-2), we obtain an expression for E. The E_z component is

$$E_z = \eta_1 \cos \theta_1 (-H'_R e^{\gamma_1 y \cos \theta_1} + H''_R e^{-\gamma_1 y \cos \theta_1}) e^{-\gamma_1 z \sin \theta_1} \qquad (2)$$

In medium 2, we have

$$H'''_x = H_R e^{\gamma_2 y \cos \theta_2} e^{-\gamma_2 z \sin \theta_2} \qquad (3)$$

$$E'''_z = -\eta_2 \cos \theta_2 H_R e^{\gamma_2 y \cos \theta_2} e^{-\gamma_2 z \sin \theta_2} \qquad (4)$$

where $H_R = H'_R + H''_R$ is the resultant magnetic intensity at the origin For polarization parallel to the plane of incidence, the effective propagation constant and characteristic wave impedance in a direction normal to the reflecting surface are

$$Z_{n1} = \eta_1 \cos \theta_1 \qquad Z_{n2} = \eta_2 \cos \theta_2 \qquad (5)$$

$$\gamma_{n1} = \gamma_1 \cos \theta_1 \qquad \gamma_{n2} = \gamma_2 \cos \theta_2 \qquad (6)$$

The reflection and transmission coefficients are again given by Eqs. (14.10-17 and 18) where the impedances are given by Eqs. (5).

The characteristic wave impedance for the wave polarized parallel to the plane of incidence is always less than the intrinsic impedance of the medium and approaches zero value as the angle of incidence approaches 90 degrees. For the wave polarized normal to the plane of incidence, the characteristic wave impedance is always greater than the intrinsic impedance and has the limiting value of ∞ as the angle of incidence approaches 90 degrees.

14.12. Oblique-incidence Reflection—Lossless Dielectric Mediums.—Many of the fundamental laws of optics may be derived from the foregoing relationships. For example, let us assume that both mediums are lossless dielectrics and derive Snell's law of refraction. We start by obtaining expressions for E_R at the boundary surface by setting $y = 0$ in Eqs.

(14.10-9 and 15). The boundary conditions require equality of tangential electric intensities on either side of the boundary; hence we have

$$E_R e^{-\gamma_1 z \sin \theta_1} = E_R e^{-\gamma_2 z \sin \theta_2} \tag{1}$$

Equating exponents and substituting $\gamma_1 = j\omega\sqrt{\mu_1\epsilon_1}$ and $\gamma_2 = j\omega\sqrt{\mu_2\epsilon_2}$, we obtain $\sqrt{\mu_1\epsilon_1/\mu_2\epsilon_2} = \sin\theta_2/\sin\theta_1$. Since the permeabilities are equal, the latter equation reduces to the familiar form of Snell's law,

$$\frac{\sin\theta_2}{\sin\theta_1} = \sqrt{\frac{\epsilon_1}{\epsilon_2}} = \frac{n_1}{n_2} \tag{2}$$

where n_1 and n_2 are the indices of refraction of mediums 1 and 2, respectively. The angle of refraction may be computed from Snell's law if the angle of incidence and the permittivities are known. Snell's law shows that the angle between the Poynting vector and the normal to the boundary surface is smallest for the medium having the highest permittivity.

A special case of interest is that in which the angle of refraction is $\theta_2 = 90$ degrees. For this case, the Poynting vector in the second medium is parallel to the boundary surface and consequently no average power crosses the boundary surface. Snell's law then becomes

$$\sin\theta_1 = \sqrt{\frac{\epsilon_2}{\epsilon_1}} \tag{3}$$

The angle θ_1 for this critical condition is known as the *angle of total internal reflection*. If the incident angle is less than the value given by Eq. (3), the propagation constant γ_{n2}, taken in a direction normal to the reflecting surface in medium 2, is imaginary and is given by $\gamma_{n2} = j\omega\sqrt{\mu_2\epsilon_2}\cos\theta_2$. The transmitted wave is then propagated without attenuation in medium 2. If, however, the angle of incidence exceeds the critical angle, the propagation constant γ_{n2} is real, indicating an exponential attenuation of the wave in medium 2. Since medium 2 is lossless, this attenuation must be due to internal reflection of the wave in the second medium.

Fresnel's equations are frequently used in geometrical optics. They express the reflection and transmission coefficients in terms of the angles of incidence and refraction. Fresnel's equations enable us to evaluate readily the intensities in the reflected and transmitted waves if the intensities of the incident wave and angles of incidence and refraction are known. The reflection and transmission coefficients for either direction of polarization are

$$r_R = \frac{Z_{n2} - Z_{n1}}{Z_{n2} + Z_{n1}} \tag{14.10-17}$$

$$r_T = \frac{2Z_{n2}}{Z_{n2} + Z_{n1}} \tag{14.10-18}$$

If the wave is polarized normal to the plane of incidence, the impedances Z_{n1} and Z_{n2} are given by Eqs. (14.10-6 and 13). For a lossless dielectric medium these become $Z_{n1} = \sqrt{\mu_1/\epsilon_1} \sec \theta_1$ and $Z_{n2} = \sqrt{\mu_2/\epsilon_2} \sec \theta_2$. Substituting these into the reflection and transmission coefficient equations and canceling out the permeabilities, we obtain

$$r_R = \frac{\sqrt{\epsilon_1/\epsilon_2} \sec \theta_2 - \sec \theta_1}{\sqrt{\epsilon_1/\epsilon_2} \sec \theta_2 + \sec \theta_1}$$

$$r_T = \frac{2\sqrt{\epsilon_1/\epsilon_2} \sec \theta_2}{\sqrt{\epsilon_1/\epsilon_2} \sec \theta_2 + \sec \theta_1}$$

Snell's law is used to eliminate $\sqrt{\epsilon_1/\epsilon_2}$, giving

$$r_R = \frac{\tan \theta_2 - \tan \theta_1}{\tan \theta_2 + \tan \theta_1} = \frac{\sin (\theta_2 - \theta_1)}{\sin (\theta_2 + \theta_1)} \tag{4}$$

$$r_T = \frac{2 \tan \theta_2}{\tan \theta_2 + \tan \theta_1} = \frac{2 \sin \theta_2 \cos \theta_1}{\sin (\theta_2 + \theta_1)} \tag{5}$$

In a similar manner r_R and r_T may be evaluated for a wave polarized parallel to the plane of incidence. The impedances Z_{n1} and Z_{n2} for this case are obtained from Eq. (14.11-5), yielding

$$r_R = -\frac{\sin 2\theta_1 - \sin 2\theta_2}{\sin 2\theta_1 + \sin 2\theta_2} = -\frac{\tan (\theta_1 - \theta_2)}{\tan (\theta_1 + \theta_2)} \tag{6}$$

$$r_T = \frac{2 \sin 2\theta_2}{\sin 2\theta_1 + \sin 2\theta_2} = \frac{\sin 2\theta_2}{\cos (\theta_1 - \theta_2) \sin (\theta_1 + \theta_2)} \tag{7}$$

Equations (4) to (7) are the Fresnel equations for dielectric mediums. In a typical problem, the known quantities might include the angle of incidence, the electric intensity of the incident wave, and the dielectric constants of the mediums. Snell's law may then be used to compute the angle of refraction θ_2. The reflection and transmission coefficients may then be computed either by the Fresnel equations or by Eqs. (14.10-17 and 18). Having the values of r_R and r_T, these are used to compute the values of electric intensities of the reflected and transmitted waves. We are then in a position to evaluate the magnetic intensities and power flow in the incident, reflected, and transmitted waves.

If the reflection coefficient has zero value, the wave is transmitted without reflection. Inspection of Eq. (6) shows that this is possible for the wave which is polarized in the plane of incidence if we have $\theta_1 + \theta_2 = 90$ degrees, since we then have $\tan (\theta_1 + \theta_2) = \infty$. The corresponding angle of incidence is known as *Brewster's angle* or the *polarizing angle*. For this

particular angle, we have $\theta_2 = 90° - \theta_1$; hence, Snell's law becomes

$$\frac{\sin \theta_1}{\cos \theta_1} = \tan \theta_1 = \sqrt{\frac{\epsilon_2}{\epsilon_1}}$$

Since the reflection coefficient is zero, Eq. (14.10-17) shows that the impedances are matched, or $Z_{n2} = Z_{n1}$.

In general, light radiation is polarized in all directions. This may be resolved into components polarized normal to the plane of incidence and components polarized parallel to the plane of incidence. If the angle of incidence is equal to Brewster's angle, the wave polarized parallel to the plane of incidence will be transmitted without reflection. The reflected wave will therefore consist of a wave which is polarized in a direction normal to the plane of incidence. This is a method which may be used to obtain polarized light.

Equations (4) and (6) show that the reflection coefficient is also zero if $\theta_1 = \theta_2$. This simply means that the two mediums have identical electrical properties, and obviously there will be no reflection under these conditions.

14.13. Wavelength and Velocity.—In our discussion of wave guides, we shall encounter several different types of wavelengths and wave velocities. It is therefore advisable to clarify these concepts by taking advantage of the simple illustrations provided by plane-wave reflection.

Consider a uniform plane wave traveling in a dielectric medium and impinging upon the surface of a perfect conductor at oblique incidence. The incident wave is shown in Fig. 6, the reflected wave being omitted for clarity. The wavelength may be defined as the distance between two successive equiphase points on the wave at any instant of time. Applying this definition to the incident wave in Fig. 6, we discover that there are many different ways in which the wavelength can be taken. For example, λ is the wavelength taken in a direction normal to the wavefront, or in the direction of propagation of the incident wave; λ_n is in a direction normal to the reflecting plane; and λ_p is parallel to the reflecting plane. Consequently, we must be careful to specify not only the magnitude of the wavelength, but also its direction with respect to the direction of travel of the wave. If θ is the angle of incidence, Fig. 6 shows that these wavelengths are related by

$$\lambda_n = \frac{\lambda}{\cos \theta} \tag{1}$$

$$\lambda_p = \frac{\lambda}{\sin \theta} \tag{2}$$

If we were asked to choose a single wavelength to represent the wave, we would in all probability select the wavelength λ measured in a direction

normal to the plane of the wavefront.   However, we shall find later in our studies of wave guides that this is the one wavelength which we cannot conveniently measure in the guide.   On the other hand, it is a simple matter to set up standing waves which enable us to measure either λ_n or λ_p. Consequently, in the wave-guide theory we shall be largely concerned with the wavelengths λ_n and λ_p.
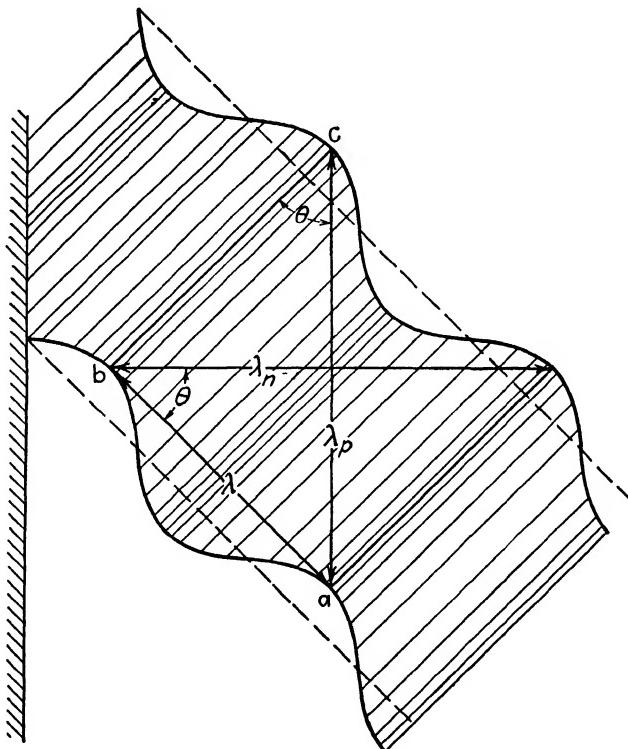

Fig. 6.—Incident wave, showing wavelengths parallel to and normal to the reflecting surface

The phase velocity may be defined for sinusoidally varying fields by the relationship $v = f\lambda$.   Any one of the three wavelengths may be used to determine a corresponding phase velocity.   The one which is commonly referred to as the phase velocity in wave guides is that which is parallel to the reflecting surface, given by $v_p = f\lambda_p$ or

$$v_p = \frac{f\lambda}{\sin \theta} = \frac{v_c}{\sin \theta} \tag{3}$$

where v_c is the velocity of light in the particular dielectric medium.   Equation (3) shows that v_p may exceed the velocity of light—in fact, it approaches infinity as θ approaches 0 radians.   It would appear that this

is inconsistent with the principle of relativity which states that the velocity of light represents the maximum attainable velocity. Referring to Fig. 6, we note that the velocity v_c may be determined by observing the time that it takes a wave particle to travel from a to b. The distance ab divided by time of travel gives the phase velocity. If we had computed the velocity using the distance ac instead of ab, the time remaining the same, we would have obtained the velocity v_p, which is greater than v_c since $ac > ab$. However, we might object to this procedure on the grounds that the wave particle actually travels from a to b and not from a to c. Thus, while the velocity v_p represents the velocity of the peak of the wave taken in the direction parallel to the reflecting surface, it does not represent the velocity of any one wave particle. Therefore the phase velocity v_p is a somewhat fictitious velocity.

14.14. Group Velocity.—The discussion thus far has been restricted to waves having sinusoidal time variation with steady-state conditions. The concept of wave velocity must be modified when dealing with nonsinusoidal waves, such as those encountered in transients or in amplitude-modulated carrier waves. In general, a modulated wave may be represented by carrier and sideband frequencies. If the wave travels in a lossless medium, all of the frequency components have the same velocity and there is no attenuation. Consequently, the wave retains its original waveform. However, if the medium is not lossless, the various frequency components have different velocities and rates of attenuation, and the wave changes its shape as it travels through the medium. We then refer to the *group velocity*, this being the velocity of a narrow band of frequencies.

Consider an amplitude-modulated wave of the form

$$E = E_0(1 + m \sin \Delta\omega t) \sin \omega t \tag{1}$$

$$= E_0 \sin \omega t + \frac{mE_0}{2} [\cos (\omega - \Delta\omega)t - \cos (\omega + \Delta\omega)t] \tag{2}$$

where $\Delta\omega$ is the angular frequency of the modulation. Equation (2) is the familiar expression containing the carrier and two side bands. It is assumed that $\omega \gg \Delta\omega$, so that the carrier and side-band frequencies occupy a narrow frequency band.

Now consider the carrier and side bands as traveling waves, the carrier having a phase constant β and the upper and lower side bands having slightly different phase constants $\beta + \Delta\beta$ and $\beta - \Delta\beta$, respectively. Writing Eq. (2) as waves traveling in the z direction, we have

$$E = E_0 \sin (\omega t - \beta z) + \frac{mE_0}{2} \{\cos [(\omega - \Delta\omega)t - (\beta - \Delta\beta)z)]$$

$$- \cos [(\omega + \Delta\omega)t - (\beta + \Delta\beta)z]\} \tag{3}$$

which may also be written

$$E = E_0[1 + m \sin (\Delta\omega t - \Delta\beta z)] \sin (\omega t - \beta z) \qquad (4)$$

The term $[1 + m \sin (\Delta\omega t - \Delta\beta z)]$ is the envelope of the modulated wave. If we were to travel along with a fixed point on the envelope, it would be necessary to satisfy the condition $\Delta\omega t - \Delta\beta z = $ a constant. Choosing the constant as zero, we have

$$\Delta\omega t - \Delta\beta z = 0 \qquad (5)$$

Our velocity would then be the group velocity

$$v_g = \frac{z}{t} = \frac{\Delta\omega}{\Delta\beta} \qquad (6)$$

or in the limit,

$$v_g = \frac{d\omega}{d\beta} = \frac{1}{d\beta/d\omega} \qquad (7)$$

The second form of Eq. (7) is usually easier to evaluate. A plane wave traveling through a lossless unbounded medium has a phase constant $\beta = \omega\sqrt{\mu\epsilon}$, and the group velocity from Eq. (7) therefore becomes $v_g = 1/\sqrt{\mu\epsilon} = v_c$. Hence, in a lossless medium the group and phase velocities are both equal to the velocity of light in the medium. In a medium with losses, β is not ordinarily directly proportional to ω and the group velocity therefore differs from the phase velocity. Such a medium is known as a dispersive medium.

We shall find, in our study of wave guides, that it is possible for the group and phase velocities to differ even though the medium may be *non-dispersive* or lossless. This occurs when the group and phase velocities are taken in some direction other than the direction of travel of the incident or reflected waves. This may be shown in a somewhat superficial way by referring to the wave in Fig. 6. Assume that the medium is lossless and that the wave particle at point a has a velocity v_c in the direction of travel of the incident wave. The component of this velocity parallel to the reflecting wall is analogous to the group velocity in wave guides, this being

$$v_g = v_c \sin \theta \qquad (8)$$

Combining with Eq. (14.13-3), we have, for a lossless medium

$$v_g v_p = v_c^2$$

Thus, if v_p and v_g are not taken in the direction of travel of the wave, they are unequal even though the medium may be lossless.

PROBLEMS

1. Compute the following for sea water at a frequency of 3,000 megacycles per sec. Constants of sea water are $\epsilon_r = 4$, $\sigma/\omega\epsilon = 0.29$.
   (a) Intrinsic impedance.
   (b) Attenuation constant.
   (c) Phase constant.
   (d) Wavelength.
   (e) Phase velocity.
   (f) Depth of penetration.
2. Repeat Prob. 1 for polystyrene.
3. A uniform plane wave is normally incident upon a body of sea water. The electric intensity of the incident wave is 100 microvolts per m and the frequency is 3,000 megacycles per sec. Compute the following:
   (a) The electric and magnetic intensities of the incident, reflected, and transmitted waves.
   (b) The time-average power densities in the incident, reflected, and transmitted waves.
4. Polystyrene has a dielectric constant of approximately 2.5. Compute the reflection coefficient and transmission coefficient for a normally incident wave passing from air into polystyrene. What would be the electrical characteristics and thickness of a material which would serve as a quarter-wavelength transformer to obtain total transmission of the incident wave at a frequency of 1,000 megacycles per sec?
5. The electric intensity of a wave in space propagating spherically outward from a source is given by

$$E_\theta = \frac{E_0 \sin \theta}{r} e^{j\omega t - \gamma r}$$

Derive the expressions for:
   (a) The magnetic intensity.
   (b) The ratio of electric to magnetic intensity.
   (c) Poynting's vector.
   (d) Using $\nabla \cdot \bar{D} = 0$, show that this field is an approximation which is valid only at very large distances from the source. (The above intensity corresponds to the radiation field of an incremental antenna, Sec. 19.02.)
6. A uniform plane wave in space has its electric intensity polarized in the y direction and is given by $E_y = E_0 e^{\gamma(lx + nz)}$. Derive expressions for the magnetic intensity and Poynting's vector.
7. A uniform plane wave in air impinges upon a body of polystyrene at an angle of incidence of 30°. The incident wave has an intensity of 100 microvolts per m and is polarized normal to the plane of incidence.
Compute:
   (a) The angle of refraction.
   (b) The electric and magnetic intensities of the incident, reflected, and refracted waves.
   (c) The power densities in the incident, reflected, and refracted waves.
   (d) Compute the normal components of power density and show that the difference between the normal components of incident- and reflected-wave power density is equal to the normal component of transmitted power density. Is this statement true for the resultant power densities?
8. A uniform plane wave in space is incident upon a body of dielectric having $\epsilon_r = 5$. The wave is polarized parallel to the plane of incidence and has an incident angle

of 45°. The electric intensity of the incident wave is 0.1 volt per m and the frequency is 100 megacycles per sec. Compute the following:

(a) The angle of refraction.

(b) The reflection and transmission coefficients by the impedance method. Check these values using Fresnel's equations.

(c) The electric and magnetic intensities of the incident, reflected, and transmitted waves.

(d) Power density in the incident, reflected, and transmitted waves.

9. Find Brewster's angle and the angle of total internal reflection for a boundary between free space and a dielectric, assuming that the dielectric has a relative permittivity of 5.

10. A uniform plane wave, polarized normal to the plane of incidence, impinges upon a copper surface with an incident angle of 50°. The electric intensity of the incident wave is 100 microvolts per m and the frequency is 1,000 megacycles per sec.

(a) Compute the electric and magnetic intensities of the incident, reflected, and transmitted waves.

(b) Compute the power densities in the incident, reflected, and transmitted waves.

(c) At what depth will the intensities in the copper have a value of $1/e$ of the surface value?

(d) Discuss the phase relationships of the incident, reflected, and transmitted waves at the surface.

11. A uniform plane wave is obliquely incident upon a plane dielectric surface.

(a) Show that if the angle of incidence is equal to or exceeds the angle of total internal reflection, the normal component of time-average power density in the second medium is zero.

(b) Will the relationship $\eta_2 = E'''/H'''$ still apply in medium 2?

(c) What are the phase relationships between the electric and magnetic intensities in medium 2 for the normal wave and for the resultant wave?

12. At high frequencies the skin-effect resistance may be found from the relationship $P = \frac{1}{2}I_m^2 R$, where P is the time-average power dissipated in the conductor, I_m is the peak value of current, and R is the skin-effect resistance. Consider a single conductor of radius a carrying a current I_m. Using Ampère's law, obtain an expression for the magnetic intensity at the surface of the conductor. Integrate the normal component of Poynting's vector over the surface of unit length of conductor to obtain the power loss and obtain an expression for the skin-effect resistance. Verify the equation for the skin-effect resistance of a coaxial line in Table 1, Chap. 8.

13. Starting with the equations for the reflection coefficient and transmission coefficient, Eqs. (14.10-17 and 18), derive Fresnel's equations for a wave polarized parallel to the plane of incidence, assuming a lossless medium.

CHAPTER 15

SOLUTION OF ELECTROMAGNETIC-FIELD PROBLEMS

The general problem of the electromagnetic field might be defined as the solution of Maxwell's equations to obtain the field intensity as a function of space and time for a given physical system. In this chapter we shall obtain the general solution of the wave equation in rectangular, cylindrical, and spherical coordinates. The general solution is a mathematical expression for the various types of waves which may exist, as referred to a given coordinate system. This solution contains a number of arbitrary constants which are evaluated in such a manner as to make the field satisfy the boundary conditions for the given physical system.

A given field distribution may be expressed in any desired system of coordinates. However, the problem of satisfying the boundary conditions is greatly simplified by the choice of a coordinate system which best suits the boundaries of the particular physical system. For example, the field distribution inside of a rectangular wave guide is best expressed in rectangular coordinates, whereas the field of a guide having circular cross section should be expressed in cylindrical coordinates.

Electromagnetic fields are produced by charges and currents. In certain types of problems, such as the determination of the radiation field of an antenna, it is necessary to express the field in terms of the currents or charges producing the field. In problems of this type, it is convenient to use scalar and vector potentials since these can be expressed in terms of the charges and currents. The electric and magnetic intensities are then obtained from the potential functions.

15.01. Scalar and Vector Potentials for Stationary Fields.—In Chap. 2, Poisson's equation was derived for the electrostatic field by inserting $\bar{D} = \epsilon \bar{E} = -\epsilon \nabla V$ into the divergence equation $\nabla \cdot \bar{D} = q_\tau$, yielding

$$\nabla^2 V = -\frac{q_\tau}{\epsilon} \qquad (2.05\text{-}8)$$

For regions in which the space-charge density is zero, Poisson's equation reduces to Laplace's equation,

$$\nabla^2 V = 0 \qquad (1)$$

The expression for $\nabla^2 V$ in spherical coordinates is given in Appendix III. If the potential V is a function of r only, Laplace's equation becomes:

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{dV}{dr} \right) = 0 \tag{2}$$

This equation has a solution of the form $V = C_1/r + C_2$. Assume that the field results from a point charge q situated at $r = 0$. If we assume that the potential is zero at $r = \infty$, we have $C_2 = 0$, and the potential is given by Eq. (2.02-7), thus

$$V = \frac{q}{4\pi \epsilon r} \tag{3}$$

Since V is a scalar function, it follows that the potential at a given point due to n discrete charges is

$$V = \sum_{i=1}^{i=n} \frac{q_i}{4\pi \epsilon r_i} \tag{4}$$

For a volume distribution of charge the potential becomes

$$V = \frac{1}{4\pi \epsilon} \int_\tau \frac{q_\tau}{r} \, d\tau \tag{5}$$

where q_τ is the charge density and r is the distance from $d\tau$ to the point where the potential is evaluated.

Equation (5) represents a solution of either Poisson's equation or Laplace's equation.[1] If this equation is integrated throughout all of space so as to include the charges everywhere, it would yield the potential at any given point. Such an integration would obviously be impractical, and so we confine our attention to a given region. The general solution of Poisson's equation then consists of the sum of the potential given by Eq. (5) (integrated over the given region) plus the solutions of Laplace's equation for the given region. The solutions of Laplace's equation may be viewed as the potentials which could be produced by charges either outside the given region or on its boundary surface. For the electrostatic field the electric intensity is related to the potential by

$$\bar{E} = -\nabla V \tag{2.03-3}$$

We now ask, is it possible to represent stationary magnetic fields by a potential similar to the electrostatic potential? Since magnetic fields are set up by currents which have direction as well as magnitude, a vector potential is needed. In Sec. 13.03 it was shown that, if a field has zero

[1] STRATTON, J. A., "Electromagnetic Theory," pp. 166–169, 192–194, McGraw-Hill Book Company, Inc., New York, 1941.

divergence, the field may be represented as the curl of a vector quantity. Equation (13.04-11) gives $\nabla \cdot \bar{B} = 0$; hence we may let

$$\bar{B} = \nabla \times \bar{A} \tag{6}$$

where $\bar{A}$ is the *vector potential*.

Writing the curl equation (13.04-7) for stationary fields, we have $\nabla \times \bar{H} = J_c$, where J_c is either the convection- or conduction-current density. Inserting $\bar{H} = \bar{B}/\mu = (1/\mu)\nabla \times \bar{A}$ from Eq. (6) into this expression, we obtain

$$\nabla \times \nabla \times \bar{A} = \mu J_c \tag{7}$$

We now use the identity $\nabla \times \nabla \times \bar{A} = \nabla(\nabla \cdot \bar{A}) - \nabla^2 \bar{A}$. The curl and divergence of a vector may be defined independently.[1] Equation (6) defined the curl of $\bar{A}$. We now let $\nabla \cdot \bar{A} = 0$, yielding $\nabla \times \nabla \times \bar{A} = -\nabla^2 \bar{A}$. Equation (7) then becomes

$$\nabla^2 \bar{A} = -\mu J_c \tag{8}$$

This equation is analogous to Poisson's equation for electrostatic fields and its solution is

$$\bar{A} = \frac{\mu}{4\pi} \int_\tau \frac{J_c}{r} \, d\tau \tag{9}$$

where r is the distance from $d\tau$ to the point at which $\bar{A}$ is evaluated and J_c is the current density at $d\tau$.

The vector potential has the same direction as the current producing it. If the current density distribution is known, Eq. (9) may be used to evaluate the vector potential. Then the magnetic intensity is obtained from Eq. (6), using the relationship $\bar{H} = \bar{B}/\mu$.

If a problem under consideration involves a finite region, then the solution of Eq. (8) is the sum of Eq. (9) (integrated over the given region) plus any solution of the equation

$$\nabla^2 \bar{A} = 0 \tag{10}$$

which satisfies the boundary conditions. The latter solution accounts for any vector potential which may result from currents outside of the given region. The current density J_c may, for example, be the current flowing in a conductor or the current produced by a stream of electrons in a vacuum tube. In those regions where $J_c = 0$, there is no contribution to the vector potential, hence the integration of Eq. (9) need not be carried out over such regions.

15.02. Scalar and Vector Potentials in Electromagnetic Fields.—Since we are primarily interested in dynamic fields rather than in stationary fields,

---

[1] As an example, the curl and divergence of the vector $\bar{E}$ in Maxwell's equations are defined by the independent equations $\nabla \times \bar{E} = -\mu(\partial \bar{H}/\partial t)$ and $\nabla \cdot \bar{E} = q_\tau/\epsilon$.

it is necessary to extend the foregoing concepts. We would logically expect that the expressions for the scalar and vector potentials of dynamic fields would be more general than those for stationary fields, and would reduce to the stationary-field equations if the electric field is due to stationary charges or the magnetic field is due to unvarying currents. Let us therefore consider the dynamic relationships, assuming a lossless medium.

Again we let

$$\bar{B} = \nabla \times \bar{A} \tag{1}$$

and substitute this into Eq. (13.04-3) to obtain $\nabla \times \bar{E} = -\nabla \times (\partial \bar{A}/\partial t)$, which may be written

$$\nabla \times \left(\bar{E} + \frac{\partial \bar{A}}{\partial t} \right) = 0 \tag{2}$$

A solution of Eq. (2) is $\bar{E} + (\partial \bar{A}/\partial t) = 0$. Since the curl of the gradient of a scalar function is always zero, a more general solution is $\bar{E} + (\partial \bar{A}/\partial t) = -\nabla V$, or

$$\bar{E} = -\nabla V - \frac{\partial \bar{A}}{\partial t} \tag{3}$$

where V is the scalar potential and $\bar{A}$ is the vector potential. For electrostatic fields we have $\partial \bar{A}/\partial t = 0$, and Eq. (3) reduces to the familiar expression $\bar{E} = -\nabla V$.

In order to obtain an expression for the scalar potential as a function of the charges, we insert $\bar{E}$ from Eq. (3) into $\nabla \cdot \bar{E} = q_\tau/\epsilon$, obtaining

$$\nabla \cdot \bar{E} = -\frac{\partial (\nabla \cdot \bar{A})}{\partial t} - \nabla^2 V = \frac{q_\tau}{\epsilon} \tag{4}$$

The curl of $\bar{A}$ has been previously defined. However, we are at liberty to define the divergence of $\bar{A}$ in any suitable manner. We shall define $\nabla \cdot \bar{A}$ in such a way that Eq. (4) reduces to an equation similar to Eq. (13.05-4) when there is no space-charge density present, and to Poisson's equation for the electrostatic field. This requires that

$$\nabla \cdot \bar{A} = -\frac{1}{v_c^2} \frac{\partial V}{\partial t} \tag{5}$$

where $v_c = 1/\sqrt{\mu \epsilon}$ is the velocity of wave propagation. Inserting Eq. (5) into (4), we obtain

$$\nabla^2 V - \frac{1}{v_c^2} \frac{\partial^2 V}{\partial t^2} = -\frac{q_\tau}{\epsilon} \tag{6}$$

This is the dynamic form of Poisson's equation. Before discussing this relationship, let us derive a similar expression for the magnetic potential.

Starting with Eq. (13.04-7)

$$\nabla \times \bar{H} = \bar{J}_c + \epsilon \frac{\partial \bar{E}}{\partial t} \tag{13.04-7}$$

we insert $\bar{H} = (1/\mu)\nabla \times \bar{A}$ from Eq. (1). The identity $\nabla \times \nabla \times \bar{A} = \nabla(\nabla \cdot \bar{A}) - \nabla^2 \bar{A}$ then yields

$$-\nabla^2 \bar{A} + \nabla(\nabla \cdot \bar{A}) = \mu \bar{J}_c + \frac{1}{v_c^2} \frac{\partial \bar{E}}{\partial t} \tag{7}$$

The electric intensity from Eq. (3) is next inserted into Eq. (7) and the scalar potential V is eliminated in the resulting equation by the use of Eq. (5). With these substitutions Eq. (7) becomes

$$\nabla^2 \bar{A} - \frac{1}{v_c^2} \frac{\partial^2 \bar{A}}{\partial t^2} = -\mu \bar{J}_c \tag{8}$$

Equations (6) and (8) are the differential equations for the scalar and vector potentials. These equations reduce to Eqs. (2.05-8) and (15.01-8), respectively, for stationary fields, *i.e.*, when $\partial/\partial t = 0$.

For regions in which $q_\tau = 0$ and $\bar{J}_c = 0$, Eqs. (6) and (8) reduce to the wave equations

$$\nabla^2 V = \frac{1}{v_c^2} \frac{\partial^2 V}{\partial t^2} \tag{9}$$

$$\nabla^2 \bar{A} = \frac{1}{v_c^2} \frac{\partial^2 \bar{A}}{\partial t^2} \tag{10}$$

If the potential is due to a point charge or a differential current element, the solutions of Eqs. (9) and (10) may be expressed in spherical coordinates as

$$V = \frac{c}{r} f\left(t \pm \frac{r}{v_c}\right) \quad \text{and} \quad \bar{A} = \frac{\bar{c}}{r} f\left(t \pm \frac{r}{v_c}\right)$$

The term $(1/r)f[t - (r/v_c)]$ represents a wave traveling radially outward from the origin, whereas $(1/r)f[t + (r/v_c)]$ represents a wave traveling radially inward.

We would expect that the solutions of the more general expressions, Eqs. (6) and (8), would consist of traveling waves which, however, take into consideration the effects of the space charge and the currents. Furthermore, these solutions must reduce to Eqs. (15.01-5 and 9) for stationary fields. The solutions satisfying these requirements are

$$V = \frac{1}{4\pi\epsilon} \int_\tau \frac{q_\tau[t - (r/v_c)]}{r} d\tau \tag{11}$$

$$\bar{A} = \frac{\mu}{4\pi} \int_\tau \frac{\bar{J}_c[t - (r/v_c)]}{r} d\tau \tag{12}$$

Since the wave travels with a finite velocity v_c, the time required for it to travel a distance r is r/v_c. Hence the potentials at a point distant r from the source at time t are due to the values of q_τ and J_c at the source at an earlier time given by $[t - (r/v_c)]$. In Eqs. (11) and (12) the quantities $q_\tau[t - (r/v_c)]$ and $J_c[t - (r/v_c)]$ represent the charges and currents at the source at time $t - r/v_c$ while the resulting potentials are at time t. The potentials in these equations are known as *retarded potentials* because of the finite time required for the wave of potential to travel from the source to the given point.

For a time function of the form $e^{j\omega t}$, Eqs. (11) and (12) become

$$V = \frac{1}{4\pi\epsilon} \int_\tau \frac{q_\tau}{r} e^{-j\beta r} \, d\tau \tag{13}$$

$$\bar{A} = \frac{\mu}{4\pi} \int_\tau \frac{J_c}{r} e^{-j\beta r} \, d\tau \tag{14}$$

where $e^{j\omega t}$ is omitted for brevity and $\beta = \omega/v_c$ is the phase constant for a lossless dielectric medium.

Also, for the time function $e^{j\omega t}$, Eqs. (9) and (10) reduce to $\nabla^2 V = -\beta^2 V$ and $\nabla^2 \bar{A} = -\beta^2 \bar{A}$. The foregoing relationships were derived for a lossless medium. For a medium which is not lossless, it can be shown that the wave equations for the potentials are similar to Eqs. (13.06-3 and 4), *i.e.*,

$$\nabla^2 V = \gamma^2 V \tag{15}$$

$$\nabla^2 \bar{A} = \gamma^2 \bar{A} \tag{16}$$

where

$$\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)} \tag{13.06-5}$$

15.03. Methods of Solving the Wave Equations.—Let us first consider a region in which there are no charges or currents; *i.e.*, $q_\tau = 0$ and $J_c = 0$. The sources of the field are then external to the region under consideration. This region might, for example, be inside of a hollow wave guide or in free space. For a time function $e^{j\omega t}$, the wave equations for $\bar{E}$ and $\bar{H}$ are given by the wave equations, Eqs. (13.06-3 and 4), while the corresponding equations for V and $\bar{A}$ are Eqs. (15.02-15 and 16).

The wave equations are second-order *partial* differential equations. They are also linear, since the potentials (or intensities) and their derivatives occur only as first-degree terms. A second-order *ordinary* differential equation has two independent solutions, each containing an arbitrary constant. The general solution may be represented as the sum of the two solutions. For example, if ϕ_1 and ϕ_2 are the two independent solutions of a second-order differential equation, the general solution is $A\phi_1 + B\phi_2$.

We shall obtain solutions for the wave equations in rectangular, cylindrical, and spherical coordinates using a method known as *separation of variables*. In this method, the partial differential equation is broken up into three ordinary differential equations, corresponding to the three independent variables (the coordinates). Each one of the ordinary differential equations is a second-order equation and therefore has two solutions as described above. The general solution of the partial differential equation is the product of the solutions of the ordinary differential equations. Thus, in rectangular coordinates, let $X(x) = C_1\phi_1(x) + C_2\phi_2(x)$, $Y(y) = C_3\phi_3(y) + C_4\phi_4(y)$, and $Z(z) = C_5\phi_5(z) + C_6\phi_6(z)$ be the solutions of the three ordinary differential equations. The general solution of the partial differential equation is then

$$f(x,y,z) = [C_1\phi_1(x) + C_2\phi_2(x)][C_3\phi_3(y) + C_4\phi_4(y)][C_5\phi_5(z) + C_6\phi_6(z)]$$

This solution contains six arbitrary constants which must be adjusted to satisfy the boundary conditions. At first sight, this appears to be a formidable task. However, most of the problems with which we will be dealing have relatively simple boundaries and the constants can therefore be easily evaluated.

The general solution of the wave equation shows that there is a very large number of field distributions possible for a given set of boundaries. The various possible field distributions are referred to as *modes*. The actual field distribution in a given physical system may consist of a single mode or a superposition of two or more modes. Although a number of modes may be required to describe the field, it should be remembered that the field itself is single-valued, *i.e.*, there is one resultant electric intensity and one magnetic intensity at any point in the field at a given instant.

Whether or not a particular mode will exist in a given region depends, in part, upon the distribution of charges and currents at the exciting source. However, the wave equation does not relate the fields to the charge or current source and therefore we cannot expect the solution of the wave equation to tell us which modes actually do exist. The solution of the wave equation informs us of all of the modes which are physically possible within the given boundaries, but it does not specify which ones actually exist.

If we are interested in evaluating the specific field set up by a given charge or current source, it is necessary to include the source in the region under consideration. The scalar and vector potentials must then be obtained by means of Eqs. (15.02-11 and 12). The intensities can then be evaluated by inserting the potentials into Eqs. (15.02-1 and 3). This method of solution is often more difficult than the solution of the wave equation, where no attempt is made to relate the fields to the source.

In the remainder of this chapter, we shall obtain the general solution of the wave equation in various coordinate systems. The analysis of wave

guides and resonators in succeeding chapters will likewise be approached from this viewpoint. The scalar and vector potential method will be used in the analysis of radiation from antennas. The wave equations for $\bar{E}$, $\bar{H}$, V, and $\bar{A}$, Eqs. (13.06-3 and 4) and (15.02-15 and 16), all have the same form and consequently we would expect them to have the same general solution. There is a distinction, however, in that $\bar{E}$, $\bar{H}$, and $\bar{A}$ are vector functions, whereas V is a scalar function. In rectangular coordinates the general solution of the wave equation is the same for either vector or scalar functions. However, in cylindrical and spherical coordinates the vector and scalar solutions differ slightly. Although the solution of the vector wave equation would be the more useful solution, it is considerably more difficult to obtain than that of the scalar wave equation. Consequently the solution for the scalar wave equation will be obtained. The solutions of the vector wave equation will be introduced as they are used.

15.04. Solution of the Wave Equation in Rectangular Coordinates.— Consider the wave equation

$$\nabla^2 V = \gamma^2 V \qquad (15.02\text{-}15)$$

applying to a region not including the source. A time function $e^{j\omega t}$ is assumed. The medium may be conducting, semiconducting, or insulating, and the intrinsic propagation constant γ may accordingly take complex or imaginary values.

It is significant to note that the wave equation reduces to Laplace's equation for the electrostatic field $\nabla^2 V = 0$ if we set $\gamma = 0$. Hence, with this substitution, the solutions of the wave equation become valid for electrostatic-field problems.

In rectangular coordinates, the wave equation is

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = \gamma^2 V \qquad (1)$$

The variables may be separated by assuming a solution of the form

$$V = X(x)Y(y)Z(z) \qquad (2)$$

where $X(x)$ is a function only of the x coordinate, $Y(y)$ is a function only of y, and $Z(z)$ is a function only of z. Inserting Eq. (2) into (1) and dividing by XYZ, we obtain

$$\frac{1}{X}\frac{\partial^2 X}{\partial x^2} + \frac{1}{Y}\frac{\partial^2 Y}{\partial y^2} + \frac{1}{Z}\frac{\partial^2 Z}{\partial z^2} = \gamma^2 \qquad (3)$$

The X, Y, and Z functions appear in separate terms on the left-hand side of Eq. (3). Since the sum of the three terms is a constant and each term is independently variable, it follows that each term must be equal to a

constant. Therefore, we equate the first term to a_x^2, the second term to a_y^2, and the third term to a_z^2, yielding three ordinary differential equations,

$$\frac{d^2X}{dx^2} = a_x^2 X \qquad \frac{d^2Y}{dy^2} = a_y^2 Y \qquad \frac{d^2Z}{dz^2} = a_z^2 Z \tag{4}$$

where

$$a_x^2 + a_y^2 + a_z^2 = \gamma^2 \tag{5}$$

Each of the equations in (4) has a general solution of the form

$$X = C_1 e^{a_x x} + C_2 e^{-a_x x} \tag{6}$$

If a_x is imaginary, we let $a_x = j a_x'$ and the solution may be written as a trigonometric function,

$$X = C_1 \cos a_x' x + C_2 \sin a_x' x \tag{7}$$

The solution may also be expressed in terms of hyperbolic functions,

$$X = C_1 \cosh a_x x + C_2 \sinh a_x x \tag{8}$$

The differential equations for the Y and Z functions have similar solutions. The values of X, Y, and Z may be inserted into Eq. (2) to obtain the solution

$$V = (C_1 e^{a_x x} + C_2 e^{-a_x x})(C_3 e^{a_y y} + C_4 e^{-a_y y})(C_5 e^{a_z z} + C_6 e^{-a_z z}) \tag{9}$$

If any one of the a's is zero, the corresponding term in Eq. (9) is replaced by $X = C_1 x + C_2$, $Y = C_3 y + C_4$, or $Z = C_5 z + C_6$.

As an example, let us return to the uniform plane waves of Sec. 14.02. The scalar and vector wave equations have the same type solutions when expressed in rectangular coordinates. Therefore the components of electric and magnetic intensity may be represented by an equation similar to Eq. (9). In Sec. 14.02, it was assumed that the electric intensity had only an x component. It was also assumed that there was no intensity variation in the x and z directions; hence $X(x)$ and $Z(z)$ are constants. Also, we have $a_x = a_z = 0$, and Eq. (5) gives $a_y = \gamma$. Thus, the solution of the wave equation for the electric intensity is $E_x = C_3 e^{\gamma y} + C_4 e^{-\gamma y}$, or in instantaneous form, $E_x = C_3 e^{j\omega t + \gamma y} + C_4 e^{j\omega t - \gamma y}$.

15.05. Solution of the Wave Equation in Cylindrical Coordinates.—In cylindrical coordinates the wave equation becomes

$$\frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho \frac{\partial V}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 V}{\partial \phi^2} + \frac{\partial^2 V}{\partial z^2} = \gamma^2 V \tag{1}$$

To separate variables, let

$$V = R(\rho)\Phi(\phi)Z(z) \tag{2}$$

where R, Φ, and Z are functions, respectively, of the ρ, ϕ and z coordinates. Inserting Eq. (2) into (1) and dividing by $R\Phi Z$, we have

$$\frac{1}{\rho R}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial R}{\partial\rho}\right) + \frac{1}{\rho^2\Phi}\frac{\partial^2\Phi}{\partial\phi^2} + \frac{1}{Z}\frac{\partial^2 Z}{\partial z^2} = \gamma^2 \tag{3}$$

The third term is a function of z only. Again the sum of the three terms is a constant and hence the third term may be set equal to a constant a_z^2 or

$$\frac{d^2 Z}{dz^2} = a_z^2 Z \tag{4}$$

This equation has a general solution of the form given by Eqs. (15.04-6, 7, or 8).

Inserting a_z^2 for the third term of Eq. (3) and multiplying by ρ^2, we have

$$\frac{\rho}{R}\frac{\partial[\rho(\partial R/\partial\rho)]}{\partial\rho} + \frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\phi^2} + (a_z^2 - \gamma^2)\rho^2 = 0 \tag{5}$$

The second term is function of ϕ only; hence, equating this to a constant which we represent by $-\nu^2$, we have the equation for Φ

$$\frac{d^2\Phi}{d\phi^2} = -\nu^2\Phi \tag{6}$$

The solution of this equation is also of the form expressed in Eqs. (15.04-6, 7 or 8), although (15.04-7) is most commonly used.

Replacing the Φ term by $-\nu^2$ in Eq. (5) and multiplying through by R, we obtain the equation for the radial function

$$\rho\frac{d}{d\rho}\left(\rho\frac{dR}{d\rho}\right) + [(a_z^2 - \gamma^2)\rho^2 - \nu^2]R = 0 \tag{7}$$

This is a form of Bessel's equation. It may be put in standard form by letting $k^2 = (a_z^2 - \gamma^2)$ and $x = k\rho$ (note: x is a new variable and is *not* the x coordinate). Equation (7) then becomes

$$x^2\frac{d^2 R}{dx^2} + x\frac{dR}{dx} + (x^2 - \nu^2)R = 0 \tag{8}$$

Since the Bessel equation is a second-order differential equation, it has two independent solutions. The solutions [1-3] are obtained by assuming an

[1] McLachlan, N. W., "Bessel Functions for Engineers," Oxford University Press, New York, 1934.

[2] Gray, A., G. B. Mathews, and T. M. MacRobert, "Bessel Functions," The Macmillan Company, New York, 1922.

[3] Watson, G. N., "Theory of Bessel Functions," Cambridge University Press, London 1922.

infinite series solution of the form

$$R = \sum_{i=0}^{\infty} C_i x^{b+i} \tag{9}$$

The coefficients C_i and the constant b in the series are evaluated by substituting Eq. (9) into (8). Thus, the series for R is differentiated twice with respect to x and then multiplied by x^2 to obtain the first term of Eq. (8). The second and third terms are evaluated in a similar manner. The three terms are then added and coefficients of like powers of x are collected. We then have an ascending power series in x which represents the left-hand side of Eq. (8). Since the sum of the series is zero for any value of x, it follows that the coefficients of any term must be zero. In this manner, the coefficients C_i and the exponent b in Eq. (9) are evaluated in terms of one undetermined constant, which is usually the first coefficient of the series.

The two solutions of Eq. (8) obtained by this procedure are the Bessel functions of the first kind and order ν:

$$J_\nu(x) = \sum_{m=0}^{\infty} \frac{(-1)^m (x/2)^{\nu+2m}}{m! \Gamma(m + \nu + 1)} \tag{10}$$

$$J_{-\nu}(x) = \sum_{m=0}^{\infty} \frac{(-1)^m (x/2)^{-\nu+2m}}{m! \Gamma(m - \nu + 1)} \tag{11}$$

where $\Gamma(m + \nu + 1) = \Gamma(p)$ is a generalized factorial function which is defined by $\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} \, dx$. This function is known as the Gamma function.

When ν takes integer values, we replace it by $\nu = n$ and the Gamma function then becomes the factorial $\Gamma(m + n + 1) = (m + n)!$. The two solutions expressed in Eqs. (10) and (11) are then related by $J_n(x) = (-1)^n J_{-n}(x)$ and hence are not independent solutions. Another solution, known as the second-kind Bessel function of order n, represented by $N_n(x)$, may also be obtained. The first- and second-kind Bessel functions of integral order are

$$J_n(x) = \sum_{m=0}^{\infty} \frac{(-1)^m (x/2)^{n+2m}}{m! (m + n)!} \tag{12}$$

$$N_n(x) = \frac{J_n(x) \cos n\pi - J_{-n}(x)}{\sin n\pi} \tag{13}$$

The first- and second-kind Bessel functions are plotted for zero order in Fig. 1 and for orders zero to five in Fig. 2. All of the Bessel functions of the second kind have negative infinite values for zero argument.

The R solution is related to the Φ solution since ν appears in both Eqs. (6) and (8). In many problems Φ will be of the form $\Phi = C_3 \cos \nu\phi + C_4 \sin \nu\phi$, and will have a periodic variation with ϕ, with a period of 2π. Thus, in a circular wave guide Φ must have the same value for $\phi = \delta$
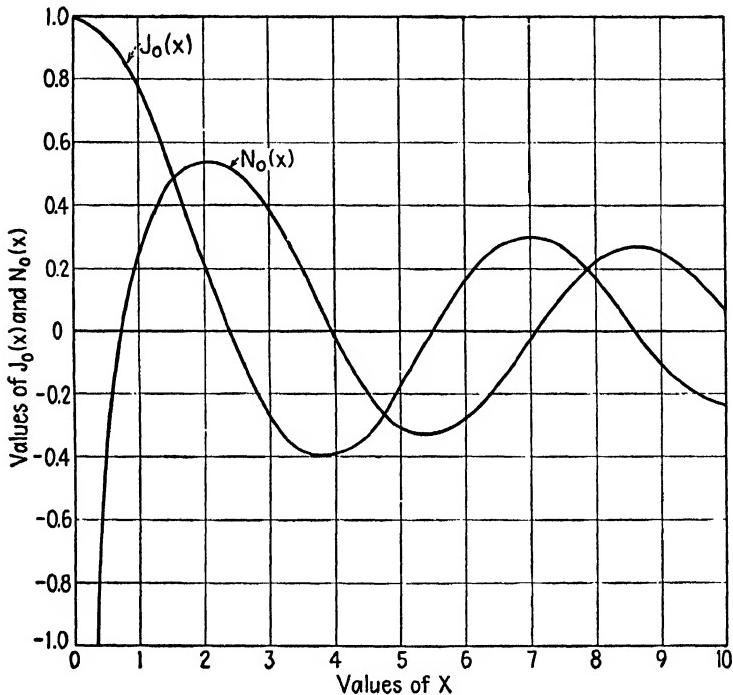
FIG. 1.—Zero-order Bessel functions of the first and second kind.

radians as it has for $\phi = \delta + 2\pi$ radians, since they correspond to the same point in the guide. This requires that ν be an integer, hence we let $\nu = n$.

The solution of the wave equation in cylindrical coordinates for integral values of $\nu = n$ is

$$V = [C_1 J_n(k\rho) + C_2 N_n(k\rho)](C_3 \cos n\phi + C_4 \sin n\phi)(C_5 e^{a_z z} + C_6 e^{-a_z z}) \quad (14)$$

while for nonintegral values of ν, we have

$$V = [C_1 J_\nu(k\rho) + C_2 J_{-\nu}(k\rho)](C_3 \cos \nu\phi + C_4 \sin \nu\phi)(C_5 e^{a_z z} + C_6 e^{-a_z z}) \quad (15)$$

Special cases occur when a_z, γ, or ν are zero. If $a_z = 0$, the Z function reduces to $Z = (C_5 z + C_6)$. If a_z and γ are both zero, then Eq. (7)
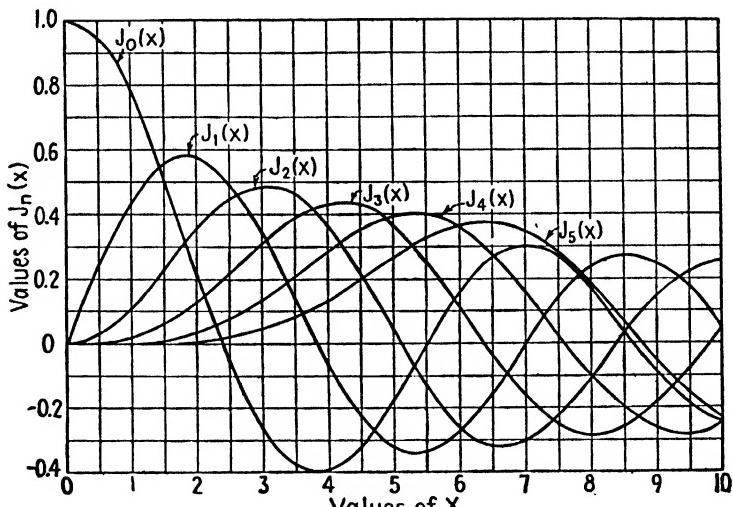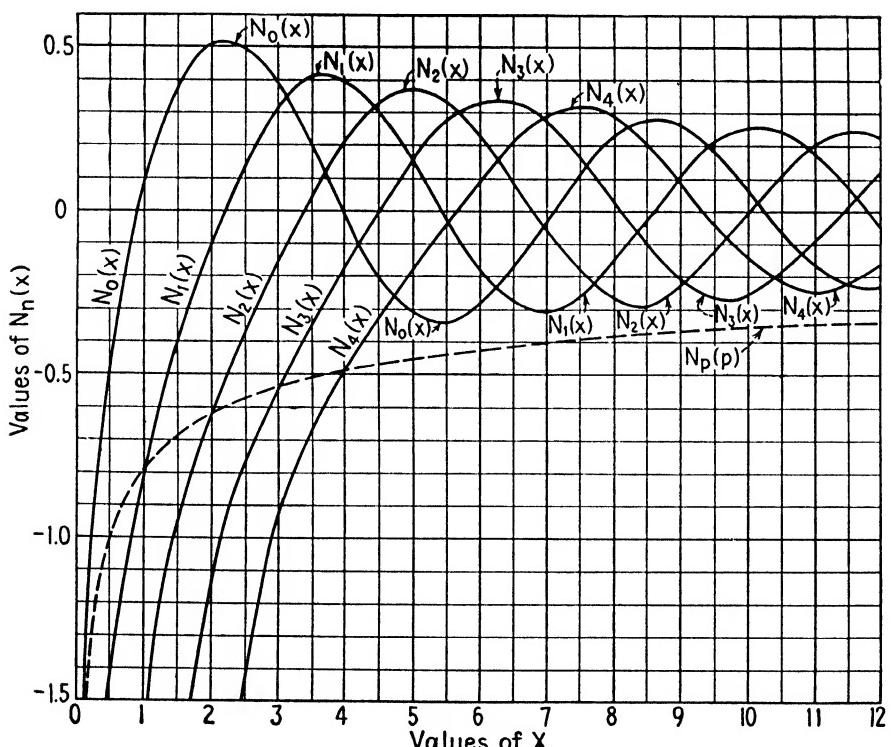
FIG. 2.—Bessel functions of the first and second kind.

becomes

$$\rho \frac{d[\rho(dR/d\rho)]}{d\rho} - \nu^2 R = 0 \tag{16}$$

which has a solution $R = (C_1\rho^\nu + C_2\rho^{-\nu})$. The potential is then

$$V = (C_1\rho^\nu + C_2\rho^{-\nu})(C_3 \cos \nu\phi + C_4 \sin \nu\phi)(C_5 z + C_6) \tag{17}$$

Finally, if a_z, γ, and ν are all zero, the R solution becomes $R = C_1 \ln \rho + C_2$ and we have

$$V = (C_1' \ln \rho + C_2)(C_3\phi + C_4)(C_5 z + C_6) \tag{18}$$

The choice of the potential equation for a particular problem depends upon the boundary conditions. For example, if the problem is an electrostatic-field problem in which the potential is known to vary linearly in the z direction, then we would have $\gamma = 0$ and $a_z = 0$, and Eq. (17) would be used. If the field also has no variation in the ϕ direction, then Eq. (18) is used, with $C_3 = 0$. The following conclusions may be drawn:

1. The general solution of the scalar wave equation is given by Eq. (14) for integer values of ν, and by Eq. (15) for noninteger values.

2. The Bessel functions of the second kind have infinite values for zero argument, hence this solution must be discarded if the field extends to the origin.

3. If the field is periodic in ϕ with a period which is a submultiple of 2π, then $\nu = n$ is an integer and the Bessel functions have integral order.

4. In a field which has no variation in the ϕ direction, we have $n = 0$. The Bessel functions are then of zero order and the ϕ function reduces to a constant.

5. If a_z, γ, or ν are zero, special cases occur which are represented by Eqs. (17) and (18).

15.06. Bessel Functions for Small and Large Arguments.—Approximate expressions may be obtained for the Bessel functions for either small or large arguments. For small arguments, *i.e.*, $x \ll 1$, the functions become

$$J_\nu(x) \approx \frac{x^\nu}{2^\nu\Gamma(\nu + 1)} \tag{1}$$

$$N_\nu(x) \approx \frac{-2^\nu\Gamma(\nu)}{\pi(x)^\nu} \tag{2}$$

For large arguments such that $x \gg 1$, the asymptotic expressions are

$$J_\nu(x) \approx \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{2\nu + 1}{4}\pi\right) \tag{3}$$

$$N_\nu(x) \approx \sqrt{\frac{2}{\pi x}} \sin\left(x - \frac{2\nu+1}{4}\pi\right) \tag{4}$$

It is interesting to observe that the Bessel functions for large arguments as shown in Fig. 2 resemble damped sinusoidal functions, and the asymptotic expressions for large arguments as given by Eqs. (3) and (4) contain cosine and sine functions.

15.07. Hankel Functions.—The Hankel functions are linear combinations of Bessel functions. There are two kinds, represented by $H_\nu^{(1)}(x)$ and $H_\nu^{(2)}(x)$, where

$$H_\nu^{(1)}(x) = J_\nu(x) + jN_\nu(x) \tag{1}$$

$$H_\nu^{(2)}(x) = J_\nu(x) - jN_\nu(x) \tag{2}$$

The Bessel functions of the second kind were found to have infinite values for $\rho = 0$. Consequently, the Hankel functions have infinite values for $\rho = 0$ and cannot represent physically realizable fields which extend to the origin. Hankel functions may be used to represent the field between the conductors of a coaxial line, since the field is discontinuous at the surface of the inner conductor and does not extend to the origin.

The asymptotic expansions for the Hankel functions for large values of x are

$$H_\nu^{(1)}(x) \approx \sqrt{\frac{2}{\pi x}}\, e^{j[x-(2\nu+1)\pi/4]} \tag{3}$$

$$H_\nu^{(2)}(x) \approx \sqrt{\frac{2}{\pi x}}\, e^{-j[x-(2\nu+1)\pi/4]} \tag{4}$$

15.08. Spherical Bessel Functions.—Bessel functions of order $n + \frac{1}{2}$ are frequently encountered in spherical waves. It is therefore convenient to have a separate representation for these functions. They are known as *spherical Bessel functions* and are represented by the lower-case letters $j_n(x)$, $n_n(x)$, $h_n^{(1)}(x)$, and $h_n^{(2)}(x)$, and are defined by [1]

$$j_n(x) = \sqrt{\frac{\pi}{2x}} J_{n+\frac{1}{2}}(x) \qquad h_n^{(1)}(x) = \sqrt{\frac{\pi}{2x}} H_{n+\frac{1}{2}}^{(1)}(x)$$

$$\tag{1}$$

$$n_n(x) = \sqrt{\frac{\pi}{2x}} N_{n+\frac{1}{2}}(x) \qquad h_n^{(2)}(x) = \sqrt{\frac{\pi}{2x}} H_{n+\frac{1}{2}}^{(2)}(x)$$

Whereas Bessel functions in general are represented by an infinite series, the spherical Bessel functions can be expressed in trigonometric form. The

---

[1] The definition of the spherical Bessel functions is that given in P. M. Morse, "Vibration and Sound," pp. 246–247, McGraw-Hill Book Company, Inc., New York, 1936.

first few may be written

$$j_0(x) = \frac{\sin x}{x} \qquad\qquad n_0(x) = -\frac{\cos x}{x}$$

$$j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x} \qquad\qquad n_1(x) = -\frac{\sin x}{x} - \frac{\cos x}{x^2} \qquad (2)$$

$$j_2(x) = \left(\frac{3}{x^3} - \frac{1}{x}\right)\sin x - \frac{3}{x^2}\cos x \qquad n_2(x) = \left(\frac{1}{x} - \frac{3}{x^3}\right)\cos x - \frac{3}{x^2}\sin x$$

15.09. Modified Bessel Functions.—In dealing with waves in a dissipative medium where γ is complex, it is often more convenient to make the substitution $k^2 = -(a_z^2 - \gamma^2)$ and $x = k\rho$ in Eq. (15.05-7). This gives the *modified Bessel equation*

$$x^2 \frac{d^2R}{dx^2} + x \frac{dR}{dx} - (x^2 + \nu^2)R = 0 \qquad (1)$$

Its solutions are the *modified Bessel functions* $I_\nu(x)$ and $I_{-\nu}(x)$, given by

$$I_\nu(x) = \sum_{m=0}^{\infty} \frac{(x/2)^{\nu+2m}}{m!\,\Gamma(\nu + m + 1)} \qquad (2)$$

$$I_{-\nu}(x) = \sum_{m=0}^{\infty} \frac{(x/2)^{-\nu+2m}}{m!\,\Gamma(-\nu + m + 1)} \qquad (3)$$

A useful linear combination of these two solutions is another solution,

$$K_\nu(x) = \frac{\pi}{2 \sin \nu\pi}[I_{-\nu}(x) - I_\nu(x)] \qquad (4)$$

For positive integral values of ν, we let $\nu = n$. For this case there results $I_{-n}(x) = I_n(x)$, necessitating a new independent solution of the form

$$K_n(x) = \frac{2}{\cos n\pi}\left(\frac{\partial I_{-n}(x)}{\partial n} - \frac{\partial I_n}{\partial n}\right) \qquad (5)$$

The relationships between the modified Bessel functions and the Bessel functions are

$$I_\nu(jx) = e^{j(\pi/2)\nu}J_\nu(x) \qquad (6)$$

$$K_\nu(jx) = \frac{\pi}{2}e^{-j(\nu+1)(\pi/2)}[J_\nu(x) - jN_\nu(x)] \qquad (7)$$

In other words, a modified Bessel function with an imaginary argument may be expressed as an ordinary Bessel function. In a dissipative medium

the propagation constant γ is complex and consequently $x = (\gamma^2 - a_z^2)^{\frac{1}{2}}\rho$ is likewise complex. The modified Bessel functions are then the most convenient expression. However, for lossless mediums, $\gamma = j\beta$ is imaginary and consequently x is imaginary. The Bessel functions are then preferred.

15.10. Other Useful Bessel-function Relationships.—In the following equations $Z_\nu(x)$ may represent any one of the functions $J_\nu(x)$, $J_{-\nu}(x)$, $N_\nu(x)$, $H_\nu^{(1)}(x)$, or $H_\nu^{(2)}(x)$.

Recurrence Formula:

$$Z_\nu(x) = \frac{x}{2\nu}[Z_{\nu-1}(x) + Z_{\nu+1}(x)] \tag{1}$$

Differentiation of Bessel Functions:

$$Z_0'(x) = -Z_1(x) \tag{2}$$

$$Z_1'(x) = Z_0(x) - \frac{1}{x}Z_1(x) \tag{3}$$

$$J_0'(x) = -J_1(x) \tag{4}$$

$$N_0'(x) = -N_1(x) \tag{5}$$

$$I_0'(x) = I_1(x) \tag{6}$$

$$K_0'(x) = -K_1(x) \tag{7}$$

$$\frac{dZ_\nu(x)}{dx} = \frac{1}{2}[Z_{\nu-1}(x) - Z_{\nu+1}(x)] \tag{8}$$

$$\frac{d[x^\nu Z_\nu(x)]}{dx} = x^\nu Z_{\nu-1}(x) \tag{9}$$

$$\frac{d[x^{-\nu}Z_\nu(x)]}{dx} = -x^{-\nu}Z_{\nu+1}(x) \tag{10}$$

Integrals of Bessel Functions:

$$\int Z_1(x)\,dx = -Z_0(x) \tag{11}$$

$$\int x^\nu Z_{\nu-1}(x)\,dx = x^\nu Z_\nu(x) \tag{12}$$

$$\int x^{-\nu}Z_{\nu+1}(x)\,dx = -x^{-\nu}Z_\nu(x) \tag{13}$$

$$\int_0^x x J_n^2(kx)\, dx = \frac{x^2}{2} \left\{ [J_n'(kx)]^2 + \left(1 - \frac{n^2}{k^2 x^2}\right) J_n^2(kx) \right\}$$

$$= \frac{x^2}{2} \left[-J_{n-1}(kx) J_{n+1}(kx) + J_n^2(kx) \right] \qquad (14)$$

$$\int_0^\infty J_n(x)\, dx = 1 \qquad (15)$$

$$\int_0^\infty \frac{J_n(kx)}{x}\, dx = \frac{1}{n} \qquad n = 1, 2, 3 \cdots \qquad (16)$$

15.11. Illustrative Example.—As an illustration of the solution of field problems in cylindrical coordinates, let us determine the potential distribution in the dielectric of the short-circuited coaxial line shown in Fig. 3
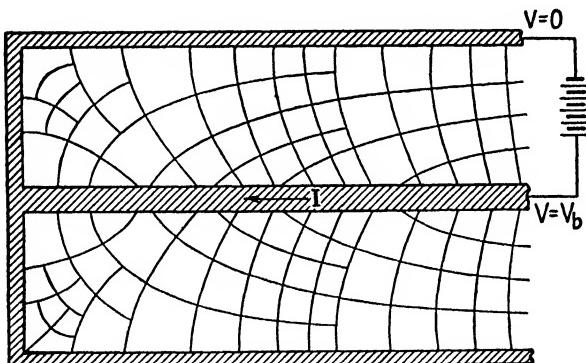
Fɪɢ. 3.—Electric field distribution in the dielectric of a coaxial line.

in the vicinity of the short-circuited end. A d-c potential difference V_b is applied between the two conductors and steady-state conditions are assumed. At the outset, let us assume that both conductors have finite conductivity.

The problem is essentially a solution of the Laplace equation $\nabla^2 V = 0$ in cylindrical coordinates. This is equivalent to setting $\gamma = 0$ in Eq. (15.05-1). The field is symmetrical in the ϕ direction; hence we have $a_\phi = n = 0$. We also have $k = \sqrt{a_z^2 - \gamma^2} = a_z$ and $x = k\rho = a_z \rho$. Thus, the potential distribution, as obtained from Eq. (15.05-14), becomes

$$V = [C_1 J_0(a_z \rho) + C_2 N_0(a_z \rho)](C_3 e^{a_z z} + C_4 e^{-a_z z}) \qquad (1)$$

The second-kind Bessel function is allowed since the field extends only to $\rho = a$, where a is the radius of the inner conductor.

If the potential varies linearly along the z axis, we have $a_z = 0$. Since

we also have $\gamma = 0$ and $n = 0$, the potential may be represented by Eq. (15.05-18) (with $C_3 = 0$), thus

$$V = (C_1 \ln \rho + C_2)(C_5 z + C_6) \tag{2}$$

The boundary conditions determine whether the solution is of the form given by Eq. (1) or (2). In the example shown in Fig. 3, the potential drop varies linearly along the Z axis, hence the potential will be given by Eq. (2). To further simplify the solution, let us assume that the outer conductor and short-circuiting end wall have infinite conductivity. If the radii of the inner and outer conductors are a and b, respectively, the boundary conditions become:

$$V = 0 \qquad \text{when} \qquad \rho = b, \qquad z = \text{any value}$$

$$V = 0 \qquad \text{when} \qquad z = 0, \qquad \rho = \text{any value}$$

$$V = V_b \qquad \text{when} \qquad \rho = a, \qquad z = l$$

The first two boundary conditions require that $C_6 = 0$ and $C_2 = -C_1 \ln b$, yielding $V = C_0 z \ln (\rho/b)$ where $C_0 = C_1 C_5$. The last boundary condition gives $C_0 = V_b/[l \ln (a/b)]$, and the final potential distribution is therefore given by

$$V = \frac{V_b z \ln (\rho/b)}{l \ln (a/b)} \tag{3}$$

The electric intensity may be obtained from the relationship $\bar{E} = -\nabla V = -(\partial V/\partial \rho)\bar{\rho} - (\partial V/\partial z)\bar{k}$. The electric field lines and equipotential lines for this case are shown in Fig. 3.

15.12. Solution of the Wave Equation in Spherical Coordinates.[1,2]—In spherical coordinates the wave equation becomes

$$\frac{1}{r^2}\frac{\partial[r^2(\partial V/\partial r)]}{\partial r} + \frac{1}{r^2 \sin \theta}\frac{\partial[\sin \theta(\partial V/\partial \theta)]}{\partial \theta} + \frac{1}{r^2 \sin^2 \theta}\frac{\partial^2 V}{\partial \phi^2} = \gamma^2 V \tag{1}$$

We will assume a solution of the form

$$V = R(r)P(\theta)\Phi(\phi) \tag{2}$$

Inserting this into Eq. (1), dividing both sides by $RP\Phi$, and multiplying by r^2, we obtain

$$\frac{1}{R}\frac{\partial[r^2(\partial R/\partial r)]}{\partial r} + \frac{1}{P \sin \theta}\frac{\partial[\sin \theta(\partial P/\partial \theta)]}{\partial \theta} + \frac{1}{\Phi \sin^2 \theta}\frac{\partial^2 \Phi}{\partial \phi^2} = \gamma^2 r^2 \tag{3}$$

[1] STRATTON, J. A., "Electromagnetic Theory," pp. 399–406, McGraw-Hill Book Company, Inc., New York, 1941.

[2] MARGENAU, H., and G. M. MURPHY, "The Mathematics of Physics and Chemistry," pp. 60–72, 216–232, D. Van Nostrand Company. Inc., New York, 1943.

The R terms are contained in the first and last terms, hence we set them equal to a_r^2, yielding

$$\frac{d[r^2(dR/dr)]}{dr} - (\gamma^2 r^2 + a_r^2)R = 0 \tag{4}$$

which may be written

$$r^2 \frac{d^2R}{dr^2} + 2r \frac{dR}{dr} - (\gamma^2 r^2 + a_r^2)R = 0 \tag{5}$$

For the moment, we drop consideration of the R solution and return to Eq. (3). Replacing the R terms by a_r^2 and multiplying through by $\sin^2 \theta$, we have

$$\frac{\sin \theta}{P} \frac{\partial[\sin \theta(\partial P/\partial \theta)]}{\partial \theta} + a_r^2 \sin^2 \theta + \frac{1}{\Phi} \frac{\partial^2 \Phi}{\partial \phi^2} = 0 \tag{6}$$

The ϕ term is equated to a constant $-m^2$, yielding

$$\frac{d^2\Phi}{d\phi^2} = -m^2\Phi \tag{7}$$

with a solution of the form $\Phi = C_5 \cos m\phi + C_6 \sin m\phi$. Again, if the field is periodic in ϕ with a period of 2π radians, m must be an integer.

Replacing the ϕ term in Eq. (6) by $-m^2$, expanding the first term, and multiplying through by $P/\sin^2 \theta$, we obtain

$$\frac{d^2P}{d\theta^2} + \cot \theta \frac{dP}{d\theta} + \left(a_r^2 - \frac{m^2}{\sin^2 \theta}\right)P = 0 \tag{8}$$

This is the *Associated Legendre equation.* Being a second-order differential equation, it has two solutions which may be obtained by the series method described for the solution of the Bessel equation. Here we find it convenient to let $a_r^2 = n(n + 1)$, where n is an integer, and $x = \cos \theta$. Equation (8) then reduces to

$$(1 - x^2) \frac{d^2P}{dx^2} - 2x \frac{dP}{dx} + \left[n(n + 1) - \frac{m^2}{1 - x^2}\right]P = 0 \tag{9}$$

The solution of Eq. (9) determines the variation of the field in the θ direction. A special case occurs when the field has circular symmetry, *i.e.*, when there is no variation in the ϕ direction. We then have $m = 0$ and Eq. (9) reduces to the *Legendre equation*

$$(1 - x^2) \frac{d^2P}{dx^2} - 2x \frac{dP}{dx} + n(n + 1)P = 0 \tag{10}$$

In the solution of Eqs. (9) and (10) by the series method, it is found that n must have integer values to avoid an infinite value of P. Furthermore,

the solutions corresponding to positive and negative values of n are related, so that we may restrict ourselves to positive integers only.   The two solutions of the Legendre equation for integral values of n are represented by $P_n(\cos \theta)$ and $Q_n(\cos \theta)$, while those of the Associated Legendre equations are $P_n^m(\cos \theta)$ and $Q_n^m(\cos \theta)$.   These are given by

$$P_n(\cos \theta) = \sum_{p=0}^{\infty} \frac{(-1)^p(n+p)!}{(n-p)!(p!)^2} \sin^{2p}\left(\frac{\theta}{2}\right) \tag{11}$$

$$Q_n(\cos \theta) = P_n(\cos \theta) \ln \cot\left(\frac{\theta}{2}\right) - \sum_{k=1}^{k=n} \frac{P_{n-k}\,P_{k-1}}{k} \tag{12}$$

$$P_n^m(\cos \theta) = (-1)^m \sin^m \theta \, \frac{d^m P_n(\cos \theta)}{d(\cos \theta)^m} \tag{13}$$

$$Q_n^m(\cos \theta) = (-1)^m \sin^m \theta \, \frac{d^m Q_n(\cos \theta)}{d(\cos \theta)^m} \tag{14}$$

The Q functions have infinite values at $\theta = 0$ and $\theta = \pi$ and hence cannot represent physically realizable fields which include these regions. Therefore the Q functions are discarded if the region under consideration includes $\theta = 0$ and $\theta = \pi$.   When n is an integer, the series given by Eq. (11) terminates after a finite number of terms and the Eqs. (11) to (14) may be represented by polynomials.   The first few Legendre and Associated Legendre polynomials are as follows:

$P_0(\cos \theta) = 1$ $P_1^1(\cos \theta) = -\sin \theta$

$P_1(\cos \theta) = \cos \theta$ $P_2^1(\cos \theta) = -3 \sin \theta \cos \theta$

$P_2(\cos \theta) = \frac{1}{2}(3 \cos^2 \theta - 1)$ $P_3^1(\cos \theta) = -\frac{3}{2} \sin \theta(5 \cos^2 \theta - 1)$

$P_3(\cos \theta) = \frac{1}{2}(5 \cos^3 \theta - 3 \cos \theta)$ $P_2^2(\cos \theta) = 3 \sin^2 \theta$ (15)

Returning now to the differential equation for R, as expressed by Eq. (5), we find that two changes of variable are required to reduce this to the standard form of the Bessel equation.   First, let $a_r^2 = n(n+1)$ in conformity with the notation used in the P-function solution, $k^2 = -\gamma^2$ and $x = kr$. Equation (5) then becomes

$$x^2 \frac{d^2R}{dx^2} + 2x \frac{dR}{dx} + [x^2 - n(n+1)]R = 0 \tag{16}$$

Now, let $R = x^{-\frac{1}{2}}W$, where W is a new variable.   Substitution into Eq. (16) gives

$$x^2 \frac{d^2W}{dx^2} + x \frac{dW}{dx} + \left[x^2 - \left(n + \frac{1}{2}\right)^2\right] W = 0 \tag{17}$$

This is the Bessel equation of order $n + \frac{1}{2}$, which has a general solution $W = C_1 J_{n+\frac{1}{2}}(x) + C_2 J_{-(n+\frac{1}{2})}(x)$. Replacing x and W by the values given above, together with a different choice of constants, yields the solution

$$R = C_1 \sqrt{\frac{\pi}{2kr}} J_{n+\frac{1}{2}}(kr) + C_2 \sqrt{\frac{\pi}{2kr}} N_{n+\frac{1}{2}}(kr)$$

$$= C_1 j_n(kr) + C_2 n_n(kr) \tag{18}$$

where $j_n(kr)$ and $n_n(kr)$ are the spherical Bessel functions defined by Eq. (15.08-1).

The R function takes a simpler form in the solution of the Laplacian equation for stationary fields. Here we set $\gamma = 0$ and $a_r^2 = n(n + 1)$ in Eq. (5), to obtain

$$r^2 \frac{d^2 R}{dr^2} + 2r \frac{dR}{dr} - n(n + 1)R = 0 \tag{19}$$

The general solution of this equation is of the form $R = Cr^a$. Substitution of this into Eq. (19) yields the values $a = n$ and $a = -(n + 1)$; hence the solution of Eq. (19) is

$$R = C_1 r^n + C_2 r^{-(n+1)} \tag{20}$$

We are now prepared to write the general solution of the wave equation in spherical coordinates. Inserting the equations for R, P, and Φ into Eq. (2), the solution is obtained:

$$V = [C_1 j_n(kr) + C_2 n_n(kr)][C_3 P_n^m(\cos\theta)$$

$$+ C_4 Q_n^m(\cos\theta)][C_5(\cos m\phi) + C_6(\sin m\phi)] \tag{21}$$

The following observations apply to this solution:

1. In stationary or quasi-stationary fields, the spherical Bessel functions are replaced by Eq. (20).

2. Integral values of n are required if infinite values of Legendre functions are to be avoided. Also, m must be an integer if the field is periodic in ϕ with a period which is a submultiple of 2π radians.

3. Since $Q_n^m(\cos\theta)$ has infinite value at $\theta = 0$ and $\theta = \pi$ radians, this solution is discarded if the field includes these regions.

4. For a circularly symmetrical field (no variation of the field in the ϕ direction), we have $m = 0$. The second term becomes the Legendre function $[C_3 P_n(\cos\theta) + C_4 Q_n(\cos\theta)]$ and the third term reduces to a constant.

15.13. Example in Spherical Coordinates.—As an example of the solution of a field problem in spherical coordinates, consider the case of a perfect dielectric sphere placed in an otherwise uniform electrostatic field. The field is assumed to be uniform at points remote from the sphere but will

be distorted in the neighborhood of the sphere. It is desired to determine the potential distribution both inside and outside the sphere.

If we choose our coordinate system such that ϕ is measured in a plane perpendicular to the field, we then have a circularly symmetrical field and consequently $m = 0$. The final bracketed term in Eq. (15.12-21) then reduces to a constant. The term $Q_n^m(\cos \theta)$ cannot be present in our solution since this would yield infinite value of potential for $\theta = 0$ and $\theta = \pi$ radians; hence we discard this solution. Since $m = 0$, the second bracketed
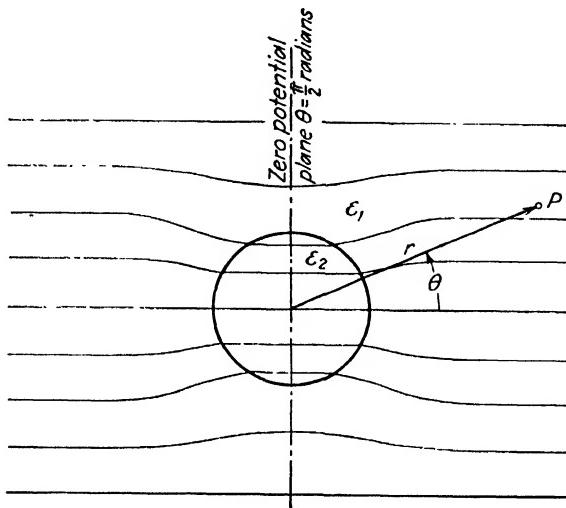
FIG. 4.—Dielectric sphere in an electrostatic field.

term becomes the Legendre function $P_n(\cos \theta)$. Also, since our problem is an electrostatic field problem, the R function is given by Eq. (15.12-20). The potential is therefore of the form

$$V = [C_1 r^n + C_2 r^{-(n+1)}]P_n(\cos \theta) \tag{1}$$

We try the solution corresponding to $n = 1$. The Legendre polynomial is $P_1(\cos \theta) = \cos \theta$, and Eq. (1) becomes

$$V_1 = \left(C_1 r + \frac{C_2}{r^2}\right) \cos \theta \tag{2}$$

In regions remote from the sphere the field is uniform; hence in Fig. 4 the potential must vary directly with distance in the horizontal direction. This distance may be taken as $r \cos \theta$. If the plane $\theta = \pi/2$ is taken as the zero potential plane, the potential in regions remote from the sphere must be $V_1 = C_1 r \cos \theta$. This is satisfied by Eq. (2) since the term C_2/r^2 is negligible at large values of r. Thus, a preliminary check shows that Eq. (2) may represent the potential distribution exterior to the sphere.

If Eq. (2) is also to represent the potential in the interior of the sphere, to avoid infinite potential at the center of the sphere it is necessary that $C_2 = 0$. Hence, in the interior of the sphere we have, as a trial solution,

$$V_2 = C_3 r \cos \theta \tag{3}$$

The constants in Eqs. (2) and (3) are evaluated by matching the solutions at the boundary of the sphere. Assuming perfectly insulating mediums, the boundary conditions require equality of tangential electric intensities and equality of normal flux densities, or

$$E_{t1} = E_{t2} \tag{4}$$

$$D_{n1} = D_{n2}$$

where subscripts 1 and 2 refer to the outer and inner mediums, respectively. We also have $E_{t1} = -(\partial V_1/r\, \partial \theta)\,|_{r=a}$ and $E_{t2} = -(\partial V_2/r\, \partial \theta)\,|_{r=a}$, where the derivatives are evaluated at the surface of the sphere; *i.e.*, at $r = a$. Equation (4) may now be written:

$$\frac{\partial V_1}{r \partial \theta}\bigg|_{r=a} = \frac{\partial V_2}{r \partial \theta}\bigg|_{r=a} \tag{5}$$

Also, $D_{n1} = \epsilon_1 E_{n1} = -\epsilon_1(\partial V_1/\partial r)\,|_{r=a}$, and similarly,

$$D_{n2} = -\epsilon_2 \frac{\partial V_2}{\partial r}\bigg|_{r=a}$$

Thus Eq. (4) becomes

$$\epsilon_1 \frac{\partial V_1}{\partial r}\bigg|_{r=a} = \epsilon_2 \frac{\partial V_2}{\partial r}\bigg|_{r=a} \tag{6}$$

Taking Eqs. (2) and (3) as the potential equations for mediums 1 and 2, respectively, and inserting these into Eqs. (5) and (6), we obtain two equations which may be solved simultaneously for the constants C_2 and C_3 in terms of C_1. This process yields

$$C_2 = -a^3 C_1 \left(\frac{\epsilon_2 - \epsilon_1}{\epsilon_2 + 2\epsilon_1} \right) \tag{7}$$

$$C_3 = \frac{3\epsilon_1 C_1}{2\epsilon_1 + \epsilon_2} \tag{8}$$

Replacing these in Eqs. (2) and (3), the potential equations become

$$V_1 = C_1 \left[r - \frac{a^3}{r^2} \left(\frac{\epsilon_2 - \epsilon_1}{\epsilon_2 + 2\epsilon_1} \right) \right] \cos \theta \tag{9}$$

$$V_2 = \frac{3C_1 \epsilon_1 r \cos \theta}{\epsilon_2 + 2\epsilon_1} \tag{10}$$

These equations satisfy all of the boundary conditions and they are solutions of the Laplace equation; hence we conclude that they represent the field distribution for the given problem.

PROBLEMS

1. A sphere of radius a contains a space charge having a density $q_r = a^2 - r^2$. Using Eq. (15.01-5), evaluate the electrostatic potential at points inside and outside the sphere. Show that these potentials satisfy Poisson's equation. Obtain expressions for the electric intensity as functions of r inside and outside the sphere.

2. Obtain the expression for the zero-order Bessel function by inserting the series given by Eq. (15.05-9) into Eq. (15.05-8) and evaluating the coefficients as described in the text. Show that the resulting series can be expressed in summed form similar to Eq. (15.05-10).

3. Prove that $\dfrac{d[x^n J_n(x)]}{dx} = x^n J_{n-1}(x)$ by inserting the series for $J_n(x)$ into the above expression and performing the indicated differentiation. Also show that

$$\frac{d[x^{-n} J_n(x)]}{dx} = -x^{-n} J_{n+1}(x)$$

4. Starting with the two identities given in Prob. 3, differentiate the left-hand sides of these equations as products and combine the resulting expressions to show that

$$\frac{dJ_n(x)}{dx} = \frac{1}{2}[J_{n-1}(x) - J_{n+1}(x)]$$

5. Obtain expressions for the potential and electric intensity in the vicinity of a conducting sphere which is placed in an otherwise uniform electrostatic field. Evaluate the space-charge density on the surface of the sphere and sketch the field.

CHAPTER 16

WAVE GUIDES

A wave guide may consist of any system of conductors or insulators which serves to guide an electromagnetic wave. A common form of wave guide consists of a hollow metallic tube containing an exciting antenna at one end and a load at the other end. The antenna sets up an electromagnetic wave which travels down the guide toward the receiving end. The conducting walls of the guide serve to confine the field and thereby to guide the electromagnetic wave. From the point of view of the Poynting theorem, the energy propagates through the dielectric inside the guide, although a small amount of the energy enters the guide walls where it is dissipated as I^2R loss.

A large number of distinct field configurations or modes are theoretically possible in wave guides. These modes correspond to solutions of Maxwell's equations which satisfy the boundary conditions of the particular guide. We shall find that a wave guide has electrical properties similar to those of a high-pass filter. A given guide has a definite cutoff wavelength for each allowed mode. If the guide is assumed to be lossless, and the wavelength of the impressed signal is shorter than the cutoff wavelength for a given mode, then electromagnetic waves can propagate down the guide in that particular mode without attenuation. However, if the wavelength of the impressed signal is appreciably longer than the cutoff wavelength, the field of the corresponding mode will be attenuated to a negligible value in a relatively short distance.

The *dominant mode* in a particular guide is the mode having the longest cutoff wavelength. It is possible to choose the dimensions of a guide in such a manner that, for a given impressed signal, only the dominant mode can be transmitted through the guide. In order for the dominant mode to exist, the width of a rectangular guide or the diameter of a circular guide must be greater than a half wavelength for the impressed signal. For this reason, wave guides are economically feasible only in the microwave portion of the frequency spectrum.

16.01. Transverse-electric (*TE*) and Transverse-magnetic (*TM*) Waves. A uniform plane wave in unbounded medium is a transverse-electromagnetic (*TEM*) wave. In such a wave the E and H vectors are both perpendicular to the direction of propagation of the wave. In wave guides, it can be

shown that the resultant wave, traveling longitudinally down the guide, can be resolved into two or more plane waves which are reflected from wall to wall in the guide as shown in Fig. 1b. This results in a component of either *E* or *H* in the direction of propagation of the resultant wave; hence the wave is not a *TEM* wave. In lossless wave guides, the modes may be classified as either *transverse electric* (*TE*) or *transverse magnetic* (*TM*) modes. In transverse electric modes, there is no component of electric intensity in the direction of propagation of the resultant wave. The electric
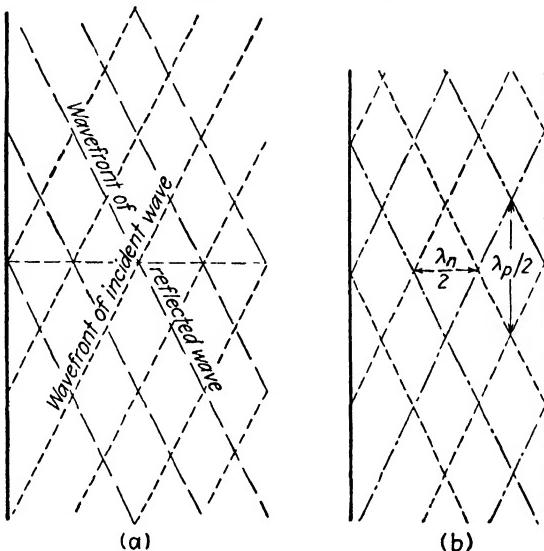
Fig. 1.—(a) Reflection of uniform plane waves at oblique incidence, and (b) reflections in a wave guide.

intensity components, therefore, lie in a plane perpendicular to the direction of propagation. In *TM* modes, there is no component of magnetic intensity in the direction of propagation. It can be shown that any wave in a lossless guide may be resolved into *TE* and *TM* components.

16.02. Wave Guides as a Reflection Phenomenon.—We may consider the fields in wave guides as being comprised of plane waves which are reflected from wall to wall in the guide in such a manner that they travel a zigzag path down the guide. The analysis of the guide, therefore, can be built up on the basis of plane-wave reflections similar to those discussed in Chap. 14. This approach offers a convenient means of visualizing the traveling waves in guides, although its application is restricted to the simpler types of modes in rectangular guides. A more rigorous approach is through the solution of Maxwell's field equations for the given boundary conditions. Let us first consider wave guides from the viewpoint of plane-wave reflections.

In Sec. 14.10, it was shown that if a uniform plane wave impinges upon a conducting surface at an angle of incidence θ, there will be a standing wave in a direction normal to the reflecting plane and a traveling wave in a direction parallel to the reflecting plane. The corresponding wavelengths are

$$\lambda_n = \frac{\lambda}{\cos \theta} \qquad (14.13\text{-}1)$$

$$\lambda_p = \frac{\lambda}{\sin \theta} \qquad (14.13\text{-}2)$$

where λ is the wavelength of the impressed signal in unbounded medium.

Let us now consider the wave polarized in a direction normal to the plane of incidence. This wave will be designated the transverse electric or *TE* wave. We will assume that the reflecting plane is perfectly conducting and that the dielectric is lossless. There will be nodal planes of electric intensity parallel to the reflecting surface and separated from it by distances

$$b = \frac{n\lambda_n}{2} \qquad (1)$$

where n is any integer. A second conducting plane may be inserted parallel to the first plane and coinciding with any one of the nodal planes without altering the field distribution between the two planes. It is assumed, of course, that the plane-wave source is situated between the two conducting planes. The addition of the second conducting plane confines the field to the region between the two planes, thereby forming a *parallel-plane* wave guide. It can be shown that the resultant wave can be resolved into two component waves, which are reflected from wall to wall of the guide as they progress down the guide.

An expression for the angle of incidence of the component waves may be obtained by eliminating λ_n from Eqs. (14.13-1) and (1), thus

$$\cos \theta = \frac{n\lambda}{2b} \qquad (2)$$

The angle of incidence, therefore, is determined by the wavelength of the impressed signal, the separation distance between the planes, and the integer n. The integer n represents the half-wave periodicity between the two planes, taken in a direction normal to the planes.

If we take the point of view that the field may be resolved into two uniform plane waves, as described above, then the angle of incidence of the component waves must satisfy Eq. (2). Any other angle of incidence would result in a tangential electric intensity at the conducting surface, which would be a violation of our boundary conditions.

The wavelength parallel to the reflecting plane is obtained by eliminating θ from Eqs. (14.13-2) and (2), yielding

$$\lambda_p = \frac{\lambda}{\sqrt{1 - (n\lambda/2b)^2}} \tag{3}$$

The cutoff wavelength λ_0 is the wavelength which makes the denominator of Eq. (3) zero or λ_p infinite, hence

$$\lambda_0 = \frac{2b}{n} \tag{4}$$

The longest cutoff wavelength occurs when $n = 1$, yielding $\lambda_0 = 2b$. Hence, for the dominant mode, the cutoff wavelength is equal to twice the separation distance between the planes.

In terms of the cutoff wavelength, the wavelength λ_p and the corresponding phase velocity v_p, both taken in a direction parallel to the reflecting walls, become

$$\lambda_p = \frac{\lambda}{\sqrt{1 - (\lambda/\lambda_0)^2}} \tag{5}$$

$$v_p = f\lambda_p = \frac{v_c}{\sqrt{1 - (\lambda/\lambda_0)^2}} \tag{6}$$

where $v_c = f\lambda$ is the velocity of light in unbounded dielectric. It should be noted that the wavelength λ is that corresponding to the impressed signal as measured in unbounded dielectric, whereas λ_n and λ_p are wavelengths in the wave guide.

Let us now consider the effect of varying the spacing between the conducting planes, assuming that the impressed wavelength λ remains constant. Equation (2) shows that as b decreases, the angle of incidence θ likewise decreases. Consequently, the wave is reflected back and forth across the guide many more times for a given distance of longitudinal travel. The wavelength λ_p and phase velocity v_p both increase as θ decreases. As the separation distance b approaches cutoff, θ approaches zero; hence the wave is reflected back and forth across the guide without any longitudinal motion. The values of v_p and λ_p both approach infinity for this limiting condition. If b is less than the cutoff value, the angle θ is imaginary and the wave is highly attenuated in the guide.

For very large values of b, the angle θ approaches 90 degrees. The wavelength λ_p and phase velocity v_p then approach the values in unbounded dielectric, λ and v_c, respectively. The wave then travels down the guide without being appreciably affected by the presence of the guide walls.

Two additional conducting planes may be added to the parallel-plane wave guide to form a closed guide as shown in Fig. 2. The electric intensity is uniform in the x direction but varies sinusoidally in the y direction, with zero values at the two side walls to satisfy the boundary conditions. Two components of magnetic intensity, H_y and H_z, are present. The resultant wave travels longitudinally down the guide.

In rectangular guides, the modes are designated $TE_{m,n}$ or $TM_{m,n}$, the integer m denoting the number of half waves of electric intensity in the x direction, while n is the number of half waves in the y direction. In the
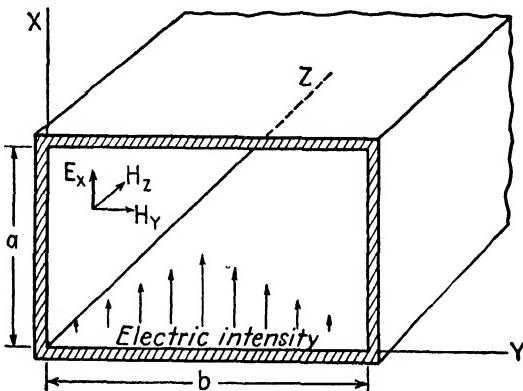
FIG. 2.—$TE_{0,n}$ mode in a rectangular guide.

mode described above, there is no variation of electric intensity in the x direction, and n half waves in the y direction; hence this corresponds to the $TE_{0,n}$ mode.

16.03. Solutions of Maxwell Equations for the $TE_{0,n}$ Mode.—Let us obtain a solution for the $TE_{0,n}$ mode using Maxwell's equations. Consider the idealized case of an infinitely long rectangular guide with perfectly conducting walls. For the present, no restrictions will be placed upon the nature of the dielectric. The guide is assumed to be excited in the $TE_{0,n}$ mode, the resultant wave intensities traveling in the z direction with a propagation term $e^{j\omega t - \Gamma z}$. It should be noted that the propagation constant Γ in the guide differs from the intrinsic propagation constant γ of the dielectric.

We start the analysis by assuming that the electric intensity is in the x direction and has a sinusoidal distribution in the y direction, thus

$$E_x = E_0 \sin \frac{n\pi y}{b} e^{j\omega t - \Gamma z} \qquad (1)$$

where b is the width of the guide and n is an integer. The electric intensity given by Eq. (1) is zero at the guide walls, $y = 0$ and $y = b$, thereby

satisfying the boundary conditions. For the assumed conditions, the two curl equations (13.06-1 and 2) reduce to

$$\frac{\partial E_x}{\partial z} = -j\omega\mu H_y \tag{2}$$

$$-\frac{\partial E_x}{\partial y} = -j\omega\mu H_z \tag{3}$$

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = (\sigma + j\omega\epsilon)E_x \tag{4}$$

Inserting E_x from Eq. (1) into (2) gives

$$\frac{E_x}{H_y} = \frac{j\omega\mu}{\Gamma} = Z_0 \tag{5}$$

where Z_0 is the *characteristic wave impedance* of the guide. Equation (5) states that the ratio of the transverse components of electric intensity to magnetic intensity is equal to a characteristic wave impedance. It will be recalled that a similar relationship was obtained for plane-wave propagation in Chap. 14.

There are two components of magnetic intensity, H_y and H_z. These may be obtained by inserting E_x from Eq. (1) into (2) and (3). For brevity, we let $k_y = n\pi/b$. The magnetic intensities then become

$$H_y = \frac{\Gamma E_0}{j\omega\mu} \sin k_y y e^{j\omega t - \Gamma z} \tag{6}$$

$$H_z = \frac{k_y E_0}{j\omega\mu} \cos k_y y e^{j\omega t - \Gamma z} \tag{7}$$

Boundary conditions require that the normal components of magnetic intensity be zero at the guide walls, a condition which is satisfied by the above equations. In fact, it can be shown that if the boundary conditions for the electric intensity are satisfied, the magnetic intensities, as determined from the curl equations, will also satisfy the boundary conditions, and vice versa.

Having determined the field intensities in the guide, we now turn to the wave equation for additional information concerning the properties of the wave in the guide. Since the electric intensity has only an x component, the wave equation (13.06-3) becomes

$$\frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} = \gamma^2 E_x \tag{8}$$

Inserting E_x from Eq. (1) into (8), we obtain

$$\Gamma^2 = \gamma^2 + k_y^2 \tag{9}$$

In this expression $\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)}$ is the intrinsic propagation constant in unbounded dielectric. In general, the propagation constant Γ is complex; hence we let

$$\Gamma = \alpha_1 + j\beta_1 \tag{10}$$

where α_1 is the attenuation constant and β_1 is the phase constant in the guide.

Consider now the special case in which the dielectric is lossless, *i.e.*, $\sigma = 0$ for the dielectric. We then have $\gamma = j\omega\sqrt{\mu\epsilon}$, and

$$\Gamma = \sqrt{k_y^2 - \omega^2\mu\epsilon} \tag{11}$$

The propagation constant Γ will be either real or imaginary, depending upon the value of ω. Cutoff occurs when $\Gamma = 0$, *i.e.*, when

$$\omega_0^2\mu\epsilon = k_y^2 \tag{12}$$

where ω_0 is the value of ω at cutoff. From the relationship $f_0\lambda_0 = v_c = 1/\sqrt{\mu\epsilon}$, we have $\omega_0^2\mu\epsilon = (2\pi/\lambda_0)^2$. Making this substitution, together with $k_y = n\pi/b$, in Eq. (12), we obtain the cutoff wavelength

$$\lambda_0 = \frac{2\pi}{k_y} = \frac{2b}{n} \tag{13}$$

which agrees with Eq. (16.02-4). For $TE_{0,n}$ modes, the cutoff wavelength is dependent upon the b dimension of the guide but is independent of the a dimension.

A convenient expression can be obtained by relating the propagation constant Γ to the impressed wavelength λ and cutoff wavelength λ_0. Inserting $\gamma^2 = -\omega^2\mu\epsilon = -(2\pi/\lambda)^2$, together with $k_y^2 = (2\pi/\lambda_0)^2$, into Eq. (9), gives us the propagation constant of the guide:

$$\Gamma = \frac{2\pi}{\lambda}\sqrt{\left(\frac{\lambda}{\lambda_0}\right)^2 - 1} \tag{14}$$

If the impressed wavelength is shorter than the cutoff wavelength, *i.e.*, if $\lambda/\lambda_0 < 1$, then Γ is imaginary and we set $\Gamma = j\beta_1$. For this condition, the wave propagates down the guide without attenuation and the intensity components undergo a phase shift of β_1 radians per unit length of guide. Conversely, if $\lambda/\lambda_0 > 1$, then Γ is real and we let $\Gamma = \alpha_1$. The intensities in the guide are then attenuated by the factor $\epsilon^{-\alpha_1 z}$.

For impressed wavelengths shorter than cutoff, we have $\Gamma = j\beta_1$, and

$$\beta_1 = \frac{2\pi}{\lambda} \sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2} \cdot \tag{15}$$

$$\lambda_p = \frac{2\pi}{\beta_1} = \frac{\lambda}{\sqrt{1 - (\lambda/\lambda_0)^2}} \tag{16}$$

$$v_p = f\lambda_p = \frac{v_c}{\sqrt{1 - (\lambda/\lambda_0)^2}} \tag{17}$$

where λ_p and v_p are again the longitudinal wavelength and phase velocity in the guide.

To obtain the group velocity, we let $\Gamma = j\beta_1$ in Eq. (11), yielding $\beta_1 = \sqrt{\omega^2\mu\epsilon - k_y^2}$. Now evaluate $d\beta_1/d\omega$ and insert this into Eq. (14.14-7) to obtain

$$v_g = v_c \sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2} \tag{18}$$

Combining Eqs. (17) and (18), we note that

$$v_g v_p = v_c^2 \tag{19}$$

Thus, the product of group velocity and phase velocity is equal to the square of the velocity of light in unbounded dielectric.

To express Z_0 in terms of λ and λ_0, substitute Γ from Eq. (14) into (5), and use $\eta = \sqrt{\mu/\epsilon}$, to obtain

$$Z_0 = \frac{\eta}{\sqrt{1 - (\lambda/\lambda_0)^2}} \tag{20}$$

If the impressed wavelength is longer than the cutoff wavelength, Γ is real and we let $\Gamma = \alpha_1$. Equation (14) then becomes

$$\alpha_1 = \frac{2\pi}{\lambda} \sqrt{\left(\frac{\lambda}{\lambda_0}\right)^2 - 1} \tag{21}$$

It is interesting to observe that the dimensionless ratios λ_p/λ, v_p/v_c, v_g/v_c, $\beta_1\lambda$, and Z_0/η at wavelengths shorter than cutoff are all functions of the wavelength ratio λ/λ_0, or the corresponding frequency ratio f_0/f. Figure 3 shows curves representing these quantities as a function of λ/λ_0 and f_0/f. These curves show that if $\lambda < \lambda_0$, then the quantities λ_p, v_p, and Z_0 all increase as λ increases, approaching infinite values as the wavelength approaches cutoff. The group velocity, on the other hand, decreases with increasing λ, approaching zero value at cutoff. Figure 3 also contains a plot of $\alpha\lambda$ against λ/λ_0, as given by Eq. (21). The attenuation is zero

when $\lambda/\lambda_0 < 1$, but rises rapidly as the ratio λ/λ_0 increases beyond unity value.

We shall find later that Èqs. (14) to (19) inclusive, as well as the curves of Fig. 3, are universally applicable to wave guides of either rectangular or circular cross section operating in any *TE* or *TM* mode.

16.04. Rectangular Guide, $TE_{m,n}$ Mode.—We now consider the more general case of $TE_{m,n}$ modes in rectangular guides. Again, it is assumed that the guide is infinitely long and has perfectly conducting walls. The
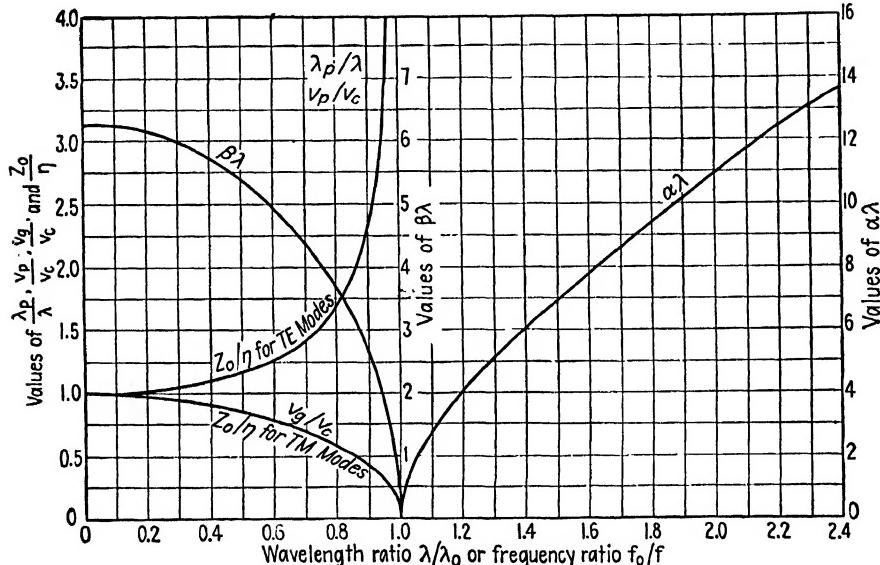
Fig. 3.—Universal curves for wave guides.

propagation term is taken as $e^{j\omega t - \Gamma z}$, denoting a wave traveling in the z direction. Since all of the intensity components contain the same propagation term, we note that, for any intensity component, we may write $\partial E/\partial z = -\Gamma E$ or $\partial H/\partial z = -\Gamma H$. Bearing this in mind, the curl equations (13.06-1 and 2) may be expanded into the following six equations.

$$\frac{\partial E_z}{\partial y} + \Gamma E_y = -j\omega\mu H_x \quad (1)$$

$$\frac{\partial H_z}{\partial y} + \Gamma H_y = (\sigma + j\omega\epsilon)E_x \quad (4)$$

$$-\Gamma E_x - \frac{\partial E_z}{\partial x} = -j\omega\mu H_y \quad (2)$$

$$-\Gamma H_x - \frac{\partial H_z}{\partial x} = (\sigma + j\omega\epsilon)E_y \quad (5)$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -j\omega\mu H_z \quad (3)$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = (\sigma + j\omega\epsilon)E_z \quad (6)$$

For transverse electric waves, we set $E_z = 0$. Equations (1) and (2) then yield

$$\frac{E_x}{H_y} = -\frac{E_y}{H_x} = \frac{j\omega\mu}{\Gamma} = Z_0 \tag{7}$$

Again we find that the ratio of the transverse components of electric intensity to magnetic intensity is equal to the characteristic wave impedance of the guide. The pair E_x and H_y result in a Poynting vector in the $+z$ direction. However, in the second pair, it is necessary to take $-E_y$
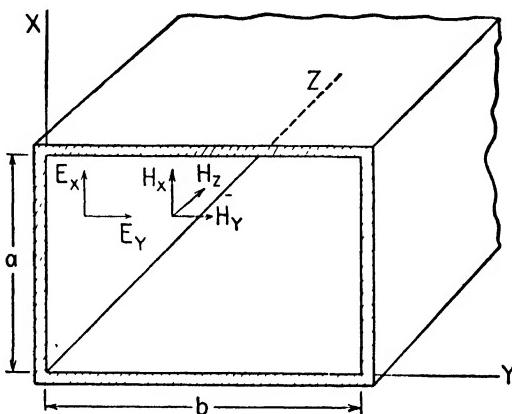
FIG. 4.—Intensity components of the $TE_{m,n}$ mode in a rectangular wave guide.

and H_x to obtain a Poynting vector in the $+z$ direction (the assumed direction of propagation of the wave). This accounts for the negative sign in the second term of Eq. (7).

In the analysis of $TE_{0,n}$ modes, we started with an assumed equation for E_x which satisfied the boundary conditions. Substitution of this value of E_x into the Maxwell equations enabled us to determine the other intensity components. However, in the analysis of the $TE_{m,n}$ modes it is easier to start with an assumed equation for the longitudinal component H_z. To be consistent with Eq. (16.03-7) and still allow for a variation of intensity in the x direction, we assume an intensity of the form

$$H_z = H_0 \cos k_x x \cos k_y y e^{j\omega t - \Gamma z} \tag{8}$$

where

$$k_x = \frac{m\pi}{a} \qquad k_y = \frac{n\pi}{b} \tag{9}$$

It will be shown later that this assumption satisfies the boundary conditions.

The wave equation (13.06-4) may be written in terms of H_z, thus

$$\nabla^2 H_z = \gamma^2 H_z \tag{10}$$

By inserting H_z from Eq. (8) into (10), we obtain

$$\Gamma^2 = \gamma^2 + k_x^2 + k_y^2 = \gamma^2 + k^2 \tag{11}$$

where $k^2 = k_x^2 + k_y^2$.   Equation (11) is similar to Eq. (16.03-9).

For a lossless dielectric, we have $\gamma = j\omega\sqrt{\mu\epsilon}$, and therefore

$$\Gamma = \sqrt{k^2 - \omega^2\mu\epsilon} \tag{12}$$

Again cutoff occurs when $\Gamma = 0$ or when $k^2 = \omega_0^2\mu\epsilon$.   Substituting $\omega_0^2\mu\epsilon = (2\pi/\lambda_0)^2$, we obtain the cutoff wavelength

$$\lambda_0 = \frac{2\pi}{k} = \frac{2\pi}{\sqrt{k_x^2 + k_y^2}} \tag{13}$$

or, by inserting the values of k_x and k_y from Eq. (9),

$$\lambda_0 = \frac{2}{\sqrt{(m/a)^2 + (n/b)^2}} \tag{14}$$

The cutoff wavelength is dependent, therefore, upon both the a and b dimensions of the guide as well as the integer values of m and n.   In general, the cutoff wavelength decreases as m and n increase, corresponding to the higher order modes.

To obtain Γ in terms of λ and λ_0, insert $\omega^2\mu\epsilon = (2\pi/\lambda)^2$ and $k^2 = (2\pi/\lambda_0)^2$ into Eq. (12), yielding

$$\Gamma = \frac{2\pi}{\lambda}\sqrt{\left(\frac{\lambda}{\lambda_0}\right)^2 - 1} \tag{15}$$

which may be inserted into Eq. (7) to obtain the characteristic wave impedance for TE modes

$$Z_0 = \frac{\eta}{\sqrt{1 - (\lambda/\lambda_0)^2}} \tag{16}$$

The propagation constant Γ for the $TE_{m,n}$ modes is identical to that obtained for $TE_{0,n}$ modes.   Consequently, the expressions for β_1, λ_p, v_p, v_g, Z_0, and α_1, as given by Eqs. (16.03-15) to (16.03-21) inclusive, and the curves in Fig. 3, are valid for all $TE_{m,n}$ modes.   The expression for the cutoff wavelength, however, is different for the two cases.   For the $TE_{m,n}$ modes, this is obtained from Eq. (14).

In order to evaluate the remaining intensities, we return to Eqs. (1) to (6) and set $E_z = 0$.   An expression for H_x is obtained by inserting F_x

from Eq. (7) into (5). Similarly, H_y is found by inserting E_x from Eq. (7) into (4). These substitutions, together with $\gamma^2 = j\omega\mu(\sigma + j\omega\epsilon)$, yield

$$H_y = \frac{\Gamma}{\gamma^2 - \Gamma^2}\frac{\partial H_z}{\partial y} \tag{17}$$

$$H_x = \frac{\Gamma}{\gamma^2 - \Gamma^2}\frac{\partial H_z}{\partial x} \tag{18}$$

(a) TE$_{0,1}$ Mode

(b) TE$_{0,2}$ Mode

(c) TE$_{1,1}$ Mode

(d) TM$_{1,1}$ Mode

—— *Electric field lines*
---- *Magnetic field lines*
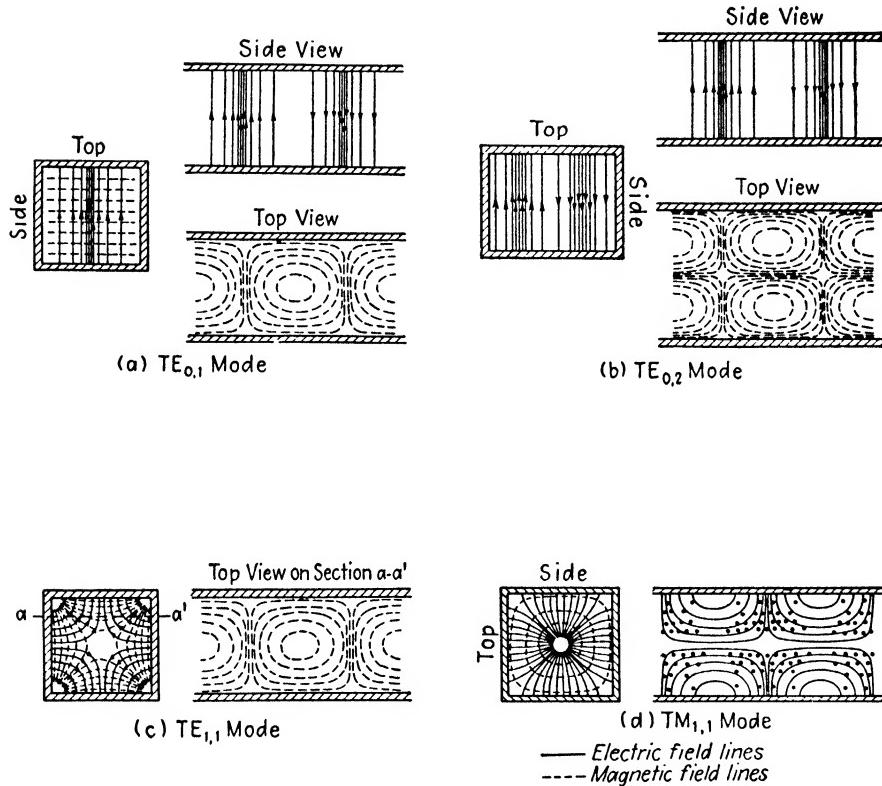
Fig. 5.—Field patterns of various modes in infinitely long rectangular guides.

The intensity H_z from Eq. (8) is now inserted into Eqs. (17) and (18). With the help of Eq. (7) and the additional relationships $\Gamma^2 - \gamma^2 = k^2 = (2\pi/\lambda_0)^2$, $\Gamma Z_0 = j\omega\mu$, and $\Gamma = j\beta_1$, the $TE_{m,n}$ mode intensities become

$$H_x = j\beta_1 k_x\left(\frac{\lambda_0}{2\pi}\right)^2 H_0 \sin k_x x \cos k_y y \tag{19}$$

$$H_y = j\beta_1 k_y\left(\frac{\lambda_0}{2\pi}\right)^2 H_0 \cos k_x x \sin k_y y \tag{20}$$

$$H_z = H_0 \cos k_x x \cos k_y y \tag{8}$$

$$E_x = H_y Z_0 = j\beta_1 k_y Z_0 \left(\frac{\lambda_0}{2\pi}\right)^2 H_0 \cos k_x x \sin k_y y \qquad (21)$$

$$E_y = -H_x Z_0 = -j\beta_1 k_x Z_0 \left(\frac{\lambda_0}{2\pi}\right)^2 H_0 \sin k_x x \cos k_y y \qquad (22)$$

$$E_z = 0 \qquad (23)$$

The propagation term $e^{j(\omega t - \beta_1 z)}$ has been omitted in the above equations for brevity.

The boundary conditions are satisfied by the above intensities, since the normal components of magnetic intensity and the tangential components of electric intensity vanish at the guide walls. The field intensity distributions for several $TE_{m,n}$ modes are shown in Fig. 5. Methods of exciting these modes are shown in Fig. 1, Chap. 18.

16.05. Rectangular Guides, $TM_{m,n}$ Modes.—The analysis of TM modes is quite similar to that of TE modes. For the TM modes, we have $H_z = 0$ and Eqs. (16.04-4) and (16.04-5) may be used to express the ratio of transverse electric to magnetic intensities, thus

$$\frac{E_x}{H_y} = -\frac{E_y}{H_x} = \frac{\Gamma}{\sigma + j\omega\epsilon} = Z_0 \qquad (1)$$

We start the analysis by assuming an axial component of electric intensity of the form

$$E_z = E_0 \sin k_x x \sin k_y y e^{j\omega t - \Gamma z} \qquad (2)$$

This satisfies the boundary conditions, since E_z is zero at the walls of the guide.

Substitution of Eq. (2) into the wave equation, written for the E_z component, $\nabla^2 E_z = \gamma^2 E_z$, yields

$$\Gamma = \sqrt{\gamma^2 + k^2} \qquad (3)$$

Equation (3) is identical to Eq. (16.04-11) for TE modes. The cutoff wavelength is obtained in the same manner as that of the TE modes and is found to be identical to Eq. (16.04-14), for a lossless guide, thus

$$\lambda_0 = \frac{2\pi}{k} = \frac{2}{\sqrt{(m/a)^2 + (n/b)^2}} \qquad (16.04\text{-}14)$$

Again we may express Γ in terms of λ and λ_0, as follows:

$$\Gamma = \frac{2\pi}{\lambda} \sqrt{\left(\frac{\lambda}{\lambda_0}\right)^2 - 1} \qquad (16.04\text{-}15)$$

Equations (16.03-15 to 19) and (16.03-21), expressing β_1, λ_p, v_p, v_g, and α_1 in terms of λ/λ_0, as well as the curves of Fig. 3, apply equally well for TM modes.

The characteristic wave impedance of *TM* modes, however, differs from that of *TE* modes. Insertion of Γ from Eq. (16.04-15) into (1) yields, for *TM* modes,

$$Z_0 = \eta \sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2} \tag{4}$$

The characteristic wave impedance of a guide is greater than the intrinsic impedance of the dielectric for *TE* modes and less than the intrinsic impedance for *TM* modes. At very short wavelengths, such that $\lambda \ll \lambda_0$, the characteristic wave impedance is approximately equal to η for both *TE* and *TM* modes. As the impressed wavelength approaches cutoff, the value of Z_0 approaches infinity for *TE* modes and zero for *TM* modes. It is interesting to note that Z_0 for *TE* and *TM* modes has the same form of frequency variation as the image impedance of π and T section high-pass filters, respectively. A comparison of the wave-guide equations with those of high-pass filters reveals many other interesting resemblances.

The remaining intensities are obtained by inserting Eq. (1) into (16.04-1 and 2), giving

$$E_y = \frac{\Gamma}{\gamma^2 - \Gamma^2} \frac{\partial E_z}{\partial y} \tag{5}$$

$$E_x = \frac{\Gamma}{\gamma^2 - \Gamma^2} \frac{\partial E_z}{\partial x} \tag{6}$$

Inserting E_z from Eq. (2) and using (1), we obtain for a guide with lossless dielectric

$$E_x = -j\beta_1 k_x \left(\frac{\lambda_0}{2\pi}\right)^2 E_0 \cos k_x x \sin k_y y \tag{7}$$

$$E_y = -j\beta_1 k_y \left(\frac{\lambda_0}{2\pi}\right)^2 E_0 \sin k_x x \cos k_y y \tag{8}$$

$$E_z = E_0 \sin k_x x \sin k_y y \tag{2}$$

$$H_x = -\frac{E_y}{Z_0} = j\frac{\beta_1}{Z_0} k_y \left(\frac{\lambda_0}{2\pi}\right)^2 E_0 \sin k_x x \cos k_y y \tag{9}$$

$$H_y = \frac{E_x}{Z_0} = -j\frac{\beta_1}{Z_0} k_x \left(\frac{\lambda_0}{2\pi}\right)^2 E_0 \cos k_x x \sin k_y y \tag{10}$$

$$H_z = 0 \tag{11}$$

[1] SHEA, T. E., "Transmission Networks and Wave Filters," pp. 228–229 D. Van Nostrand Company, Inc., New York, 1929.

If either k_x or k_y are zero, the field intensities all vanish. Hence there cannot be a $TM_{0,n}$ or $TM_{m,0}$ mode in rectangular guides comparable to the $TE_{0,n}$ mode. The $TE_{0,1}$ mode is therefore the dominant mode in rectangular wave guides.

16.06. Wave Guides of Circular Cross Section.—In the analysis of circular guides, we again assume an infinitely long guide with perfectly conducting walls and an outgoing wave with a propagation term $e^{j\omega t - \Gamma z}$.
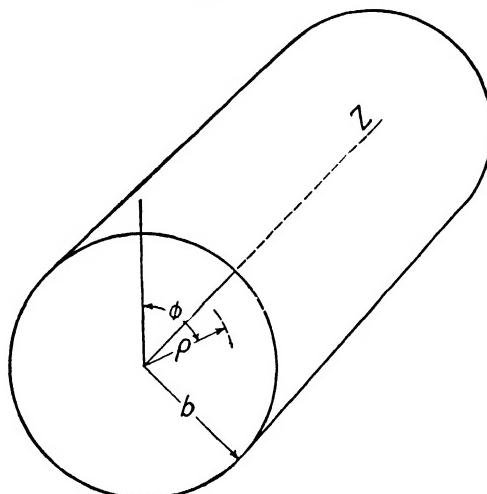
FIG. 6.—Circular wave guide.

The curl equations (13.06-1 and 2) in cylindrical coordinates may be written in the following form by expanding the determinant $\nabla \times \bar{E}$, given in Appendix III, and replacing $\partial E / \partial z$ by $-\Gamma E$:

$$\frac{1}{\rho} \frac{\partial E_z}{\partial \phi} + \Gamma E_\phi = -j\omega\mu H_\rho \tag{1}$$

$$-\Gamma E_\rho - \frac{\partial E_z}{\partial \rho} = -j\omega\mu H_\phi \tag{2}$$

$$\frac{1}{\rho} \frac{\partial(\rho E_\phi)}{\partial \rho} - \frac{1}{\rho} \frac{\partial E_\rho}{\partial \phi} = -j\omega\mu H_z \tag{3}$$

$$\frac{1}{\rho} \frac{\partial H_z}{\partial \phi} + \Gamma H_\phi = (\sigma + j\omega\epsilon) E_\rho \tag{4}$$

$$-\Gamma H_\rho - \frac{\partial H_z}{\partial \rho} = (\sigma + j\omega\epsilon) E_\phi \tag{5}$$

$$\frac{1}{\rho} \frac{\partial(\rho H_\phi)}{\partial \rho} - \frac{1}{\rho} \frac{\partial H_\rho}{\partial \phi} = (\sigma + j\omega\epsilon) E_z \tag{6}$$

For *TE* modes, we have $E_z = 0$ and for *TM* modes $H_z = 0$. Upon inserting $E_z = 0$ into Eqs. (1) and (2), and $H_z = 0$ into Eqs. (4) and (5), we again find that the ratios of transverse components of electric to magnetic intensity are equal to the characteristic wave impedances:

$$\frac{E_\rho}{H_\phi} = -\frac{E_\phi}{H_\rho} = \frac{j\omega\mu}{\Gamma} = Z_0 \qquad TE \text{ modes} \qquad (7)$$

$$\frac{E_\rho}{H_\phi} = -\frac{E_\phi}{H_\rho} = \frac{\Gamma}{\sigma + j\omega\epsilon} = Z_0 \qquad TM \text{ modes} \qquad (8)$$

The type of solution for the wave equation in cylindrical coordinates was discussed in Sec. 15.05. For an outgoing wave, the axial intensity components are of the form given by Eq. (15.05-14), thus

$$H_z = [C_1 J_n(k\rho) + C_2 N_n(k\rho)](C_3 \cos n\phi + C_4 \sin n\phi)e^{j\omega t - \Gamma z} \quad TE \text{ modes} \quad (9)$$

$$E_z = [C_1 J_n(k\rho) + C_2 N_n(k\rho)](C_3 \cos n\phi + C_4 \sin n\phi)e^{j\omega t - \Gamma z} \quad TM \text{ modes} \quad (10)$$

If the field extends to the axis, the second-kind Bessel functions must be discarded since they all have infinite values for zero arguments. The terms $\sin n\phi$ and $\cos n\phi$ determine the angular variation of the intensities. They are essentially the same type of variation since, by rotating the reference axis through an angle of $\pi/2$ radians, the sine function becomes a cosine function and vice versa. To simplify writing the intensity equations, we retain the cosine function and discard the sine function. The z-component intensities are then expressed as

$$H_z = H_0 J_n(k\rho) \cos n\phi \, e^{j\omega t - \Gamma z} \qquad TE \text{ modes} \qquad (11)$$

$$E_z = E_0 J_n(k\rho) \cos n\phi \, e^{j\omega t - \Gamma z} \qquad TM \text{ modes} \qquad (12)$$

In order to have single-valued intensities, we must choose n such that $\cos n\phi = \cos n(\phi + 2\pi)$. This requires that n have integer values, consequently we are concerned only with Bessel functions of integer orders.

To obtain expressions for the intensities, we return to the curl equations (1) to (6) and let $E_z = 0$ for *TE* modes and $H_z = 0$ for *TM* modes. For the *TE* modes, we insert E_ϕ and E_ρ from Eq. (7) into Eqs. (5) and (4), respectively. For the *TM* modes, insert H_ϕ and H_ρ from Eq. (8) into (2) and (1), respectively. We then have

$$\begin{array}{c} \underline{TE \text{ modes}} \\[6pt] \left. \begin{aligned} H_\rho &= \frac{\Gamma}{\gamma^2 - \Gamma^2} \frac{\partial H_z}{\partial \rho} \\[10pt] H_\phi &= \frac{\Gamma}{\rho(\gamma^2 - \Gamma^2)} \frac{\partial H_z}{\partial \phi} \end{aligned} \right\} (13) \end{array} \qquad \begin{array}{c} \underline{TM \text{ modes}} \\[6pt] \left. \begin{aligned} E_\rho &= \frac{\Gamma}{\gamma^2 - \Gamma^2} \frac{\partial E_z}{\partial \rho} \\[10pt] E_\phi &= \frac{\Gamma}{\rho(\gamma^2 - \Gamma^2)} \frac{\partial E_z}{\partial \phi} \end{aligned} \right\} (14) \end{array}$$

Inserting E_z and H_z from Eqs. (11) and (12) into these expressions, with the additional substitutions of Eqs. (7) and (8), we obtain the intensity equations. The propagation term $e^{j\omega t - \Gamma z}$ has been omitted for brevity and the substitutions $\Gamma = j\beta_1$, and $\Gamma^2 - \gamma^2 = k^2$ (as given by Eq. (17) below) have been made.

$$\underline{TE \text{ modes}} \qquad\qquad\qquad \underline{TM \text{ modes}}$$

$$\left.\begin{aligned}
H_\rho &= -\frac{j\beta_1}{k} H_0 J_n'(k\rho) \cos n\phi \\[2mm]
H_\phi &= \frac{j\beta_1 n}{k^2 \rho} H_0 J_n(k\rho) \sin n\phi \\[2mm]
H_z &= H_0 J_n(k\rho) \cos n\phi \\[2mm]
E_\rho &= \frac{\omega\mu}{\beta_1} H_\phi \\[2mm]
E_\phi &= -\frac{\omega\mu}{\beta_1} H_\rho \\[2mm]
E_z &= 0
\end{aligned}\right\} (15)
\qquad
\left.\begin{aligned}
E_\rho &= -\frac{j\beta_1}{k} E_0 J_n'(k\rho) \cos n\phi \\[2mm]
E_\phi &= \frac{j\beta_1 n}{k^2 \rho} E_0 J_n(k\rho) \sin n\phi \\[2mm]
E_z &= E_0 J_n(k\rho) \cos n\phi \\[2mm]
H_\rho &= -\frac{\omega\epsilon}{\beta_1} E_\phi \\[2mm]
H_\phi &= \frac{\omega\epsilon}{\beta_1} E_\rho \\[2mm]
H_z &= 0
\end{aligned}\right\} (16)$$

In order to satisfy the boundary conditions, E_z and E_ϕ must both be zero at the guide wall where $\rho = b$. For *TE* modes, this requires that $J_n'(kb) = 0$ and for *TM* modes, $J_n(kb) = 0$. The functions $J_n(kb)$ plotted against (kb) are shown in Figs. 1 and 2 in Chap. 15. There are theoretically an infinite number of discrete values of (kb) which yield $J_n'(kb) = 0$ and $J_n(kb) = 0$. For convenience in designating the modes, the successive

TABLE 5

TE modes			*TM* modes		
Roots of $J_n'(kb) = 0$			Roots of $J_n(kb) = 0$		
$TE_{0,1}$ 3.83	$TE_{0,2}$ 7.02	$TE_{0,3}$ 10.17	$TM_{0,1}$ 2.40	$TM_{0,2}$ 5.52	$TM_{0,3}$ 8.65
$TE_{1,1}$ 1.84	$TE_{1,2}$ 5.33	$TE_{1,3}$ 8.54	$TM_{1,1}$ 3.83	$TM_{1,2}$ 7.02	$TM_{1,3}$ 10.17
$TE_{2,1}$ 3.05	$TE_{2,2}$ 6.71	$TE_{2,3}$ 9.97	$TM_{2,1}$ 5.14	$TM_{2,2}$ 8.42	$TM_{2,3}$ 11.62

values of (kb) satisfying these conditions are represented by the integers $m = 1, 2, 3$, etc. Hence, the designations for circular wave-guide modes are $TE_{n,m}$ and $TM_{n,m}$ modes. From the viewpoint of the field distribution in the guide, Eqs. (11) and (12) show that n represents the number of full cycles of variation of E_z or H_z as ϕ varies through 2π radians. The

integer m represents the number of times E_ϕ is zero along a radial of the guide, the radial extending from the axis to the inner surface of the guide, the zero on the axis being excluded if it exists. Table 5 gives the first few roots of the equations $J_n(kb) = 0$ and $J'_n(kb) = 0$.

It now remains to evaluate the propagation constant and cutoff wavelength for circular guides.

The constant k in the above equations is related to the propagation constant Γ in the guide and the intrinsic propagation constant γ of the dielectric in a manner similar to that of Eq. (16.04-11). We could obtain this relationship by the procedure used for rectangular guides, that is, by inserting either H_z or E_z from Eqs. (11) or (12) into the wave equation and solving for Γ. However, this involves a little difficulty in the manipulation of Bessel functions. A simpler procedure is to return to the treatment of Bessel functions in Sec. 15.05, where we had previously set $k^2 = a_z^2 - \gamma^2$. To obtain a_z, we recall that the Z function solution in cylindrical coordinates is of the type $Z = C_1 e^{a_z z} + C_2 e^{-a_z z}$. Since we are concerned only with a wave traveling in the $+z$ direction, the first term is discarded. Comparison of the remaining term with Eqs. (11) and (12) shows that $a_z = \Gamma$. Therefore for either the *TE* or *TM* modes, the propagation constant becomes

$$k^2 = \Gamma^2 - \gamma^2$$
$$\Gamma = \sqrt{k^2 + \gamma^2} \tag{17}$$

The cutoff wavelength is obtained by writing Eq. (17) in the form

$$\frac{(kb)^2}{b^2} = \Gamma^2 - \gamma^2 \tag{18}$$

from which we obtain, for lossless dielectric,

$$\Gamma = \sqrt{\frac{(kb)^2}{b^2} - \omega^2 \mu\epsilon} \tag{19}$$

Cutoff occurs when $\Gamma = 0$, or when $\omega_0^2 \mu\epsilon = (kb)^2/b^2$. Also we have $\omega_0^2 \mu\epsilon = (2\pi/\lambda_0)^2$; hence

$$\lambda_0 = \frac{2\pi}{k} = \frac{2\pi b}{(kb)} \tag{20}$$

The values of (kb) are given in Table 5 for various modes. The $TE_{1,1}$ mode has the longest cutoff wavelength and is therefore the dominant mode in circular guides. For this mode, we have $(kb) = 1.84$ and $\lambda_0 = 3.41b$.

The propagation constant may be expressed in terms of λ and λ_0 by substituting $\omega^2 \mu\epsilon = (2\pi/\lambda)^2$ and $(kb)^2/b^2 = (2\pi/\lambda_0)^2$ into Eq. (18), giving the familiar expression

$$\Gamma = \frac{2\pi}{\lambda} \sqrt{\left(\frac{\lambda}{\lambda_0}\right)^2 - 1} \tag{21}$$

(a) TE$_{1,1}$ Mode

(b) TM$_{0,1}$ Mode

(c) TE$_{0,1}$ Mode

Section through c-d

(d) TM$_{1,1}$ Mode
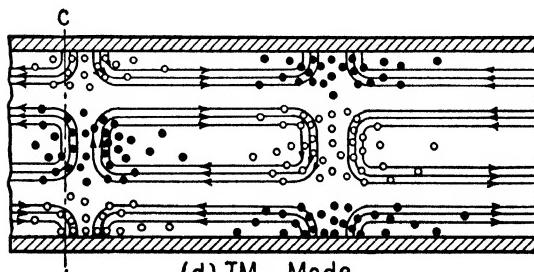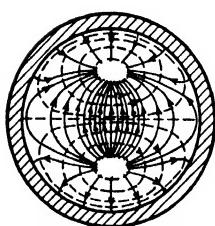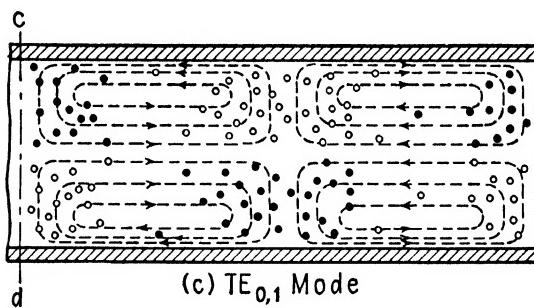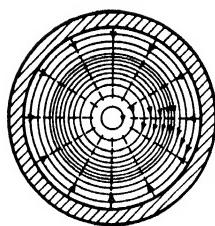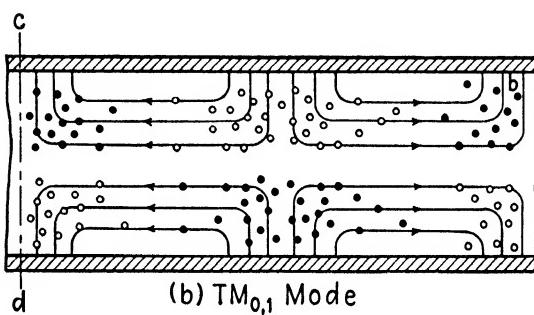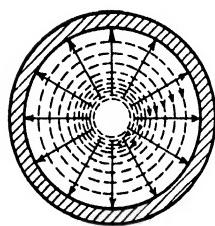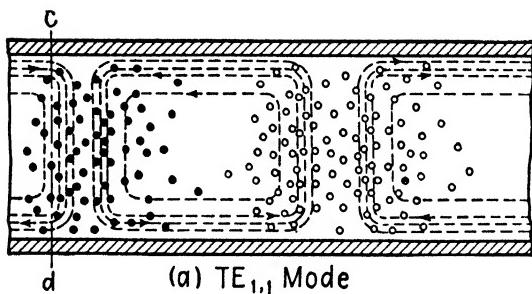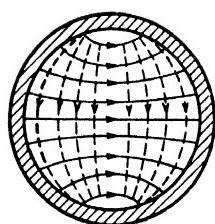
——— *Electric field lines*   − − −*Magnetic field lines*

Fig. 7.—Field patterns of various modes in circular guides.

Since this is identical to Eq. (16.03-14), it follows that Eqs. (16.03-15) to (19) and the curves in Fig. 3 expressing the values of β_1, λ_p, v_p, v_g, and α apply. The characteristic wave impedance in circular guides is given by Eq. (16.04-16) for TE modes and (16.05-4) for TM modes. The field patterns corresponding to several modes in circular guides are shown in Fig. 7.

16.07. TEM Mode in Coaxial Lines.—The principal mode or TEM mode is the most common of all modes. It has no cutoff frequency and is the mode of operation of transmission lines at low frequencies or on direct current. This mode cannot exist in wave guides since it requires two conductors, such as provided by the open-wire or coaxial transmission line.

The analysis of the TEM mode affords an excellent opportunity to illustrate the relationships between the circuit method of approach and the analysis based upon the field equations. In the following analysis it is assumed that the conductors have infinite conductivity and that the dielectric is lossless. The electric lines are then radial, terminating on charges on the conductor surfaces, and the magnetic lines are concentric circles linking the current in the center conductor. The effect of losses would be to introduce a small axial component of electric intensity and the field would no longer be a truly TEM mode, although it may closely approximate it.

In the coaxial line of Fig. 8, we have two intensities, E_ρ and H_ϕ. The curl equations (13.06-1 and 2), expressed in cylindrical coordinates, become

$$\frac{\partial E_\rho}{\partial z} = -j\omega\mu H_\phi \tag{1}$$

$$\frac{\partial H_\phi}{\partial z} = -j\omega\epsilon E_\rho \tag{2}$$

With a propagation term of the form $\epsilon^{j\omega t - \Gamma z}$, Eqs. (1) and (2) reduce to

$$\Gamma E_\rho = j\omega\mu H_\phi \tag{3}$$

$$\Gamma H_\phi = j\omega\epsilon E_\rho \tag{4}$$

from which we obtain the characteristic wave impedance

$$\frac{E_\rho}{H_\phi} = \frac{j\omega\mu}{\Gamma} = \frac{\Gamma}{j\omega\epsilon} \tag{5}$$

Solving Eq. (5) for Γ, we obtain

$$\Gamma = j\omega\sqrt{\mu\epsilon} \tag{6}$$

If this is inserted into Eq. (5), there results

$$\frac{E_\rho}{H_\phi} = \sqrt{\frac{\mu}{\epsilon}} = \eta \tag{7}$$

The propagation constant and characteristic wave impedance are therefore equal to the corresponding values in unbounded dielectric.

Ampère's law may be used to relate the magnetic intensity to the current. For a current in the center conductor given by $I = I_0 e^{j\omega t - \Gamma z}$, Ampère's law $\oint \bar{H} \cdot d\bar{l} = I$ yields $2\pi\rho H_\phi = I_0 e^{j\omega t - \Gamma z}$. Solving for H_ϕ and inserting this into Eq. (7) to obtain E_ρ, we have

$$H_\phi = \frac{I_0}{2\pi\rho} e^{j\omega t - \Gamma z} \tag{8}$$

$$E_\rho = \sqrt{\frac{\mu}{\epsilon}} \frac{I_0}{2\pi\rho} e^{j\omega t - \Gamma z} \tag{9}$$

The potential rise from the outer conductor to the center conductor is found by inserting E_ρ from Eq. (9) into $V = -\int_b^a E_\rho \, d\rho$, yielding

$$V = \frac{I_0}{2\pi} \sqrt{\frac{\mu}{\epsilon}} \ln \frac{b}{a} e^{j\omega t - \Gamma z} \tag{10}$$

The characteristic impedance of a coaxial line is equal to the ratio of voltage to current for the outgoing wave. Dividing the voltage in Eq. (10) by the current $I = I_0 e^{j\omega t - \Gamma z}$, we obtain

$$Z_0 = \frac{V}{I} = \frac{1}{2\pi} \sqrt{\frac{\mu}{\epsilon}} \ln \frac{b}{a}$$

$$= \frac{138}{\sqrt{\epsilon_r}} \log_{10} \frac{b}{a} \tag{11}$$

which agrees with the characteristic impedance given in Table 1, Chap. 8.

The curl equations (1) and (2) may be expressed in terms of voltage and current, whereupon they take on the familiar form of the differential equations of a lossless transmission line. From Eqs. (9) and (10) we obtain $E_\rho = V/[\rho \ln (b/a)]$. Also, since $I = I_0 e^{j\omega t - \Gamma z}$, we have from Eq. (8), $H_\phi = I/2\pi\rho$. Inserting these values of E_ρ and H_ϕ into Eqs. (1) and (2) gives

$$\frac{\partial V}{\partial z} = -j\omega \left(\frac{\mu}{2\pi} \ln \frac{b}{a} \right) I = -j\omega L I \tag{12}$$

$$\frac{\partial I}{\partial z} = -j\omega \left(\frac{2\pi\epsilon}{\ln (b/a)} \right) V = -j\omega C V \tag{13}$$

These are the transmission-line equations for a lossless line. The bracketed term in Eq. (12) is the inductance per unit length of line, while the bracketed term in Eq. (13) is the capacitance per unit length.

By integrating Poynting's vector over the cross section of the dielectric, it may be shown that the power is transmitted through the dielectric. Equation (14.04-1) gives the time-average power through unit area of dielectric as $\mathcal{P}_T = \frac{1}{2}E_\rho H_\phi$. The direction of power flow is mutually perpendicular to both E_ρ and H_ϕ and hence is in the z direction. Dropping the term $\epsilon^{j\omega t - \Gamma z}$ in Eqs. (8) and (9), we have, for the amplitudes, $H_\phi = I_0/2\pi\rho$ and $E_\rho = \sqrt{\mu/\epsilon}(I_0/2\pi\rho)$. The time-average power density is therefore $\mathcal{P}_T = (I_0^2/8\pi^2\rho^2)\sqrt{\mu/\epsilon}$. The total power transmitted through the dielectric is found by integrating the power density over the cross section of the dielectric, or

$$P_T = \int_a^b \mathcal{P}_T 2\pi\rho \, d\rho$$

$$= \frac{I_0^2}{4\pi}\sqrt{\frac{\mu}{\epsilon}} \ln \frac{b}{a} \tag{14}$$

Since the voltage amplitude from Eq. (10) is $V_0 = (I_0/2\pi)\sqrt{\mu/\epsilon} \ln (b/a)$, Eq. (14) may be expressed as

$$P_T = \frac{1}{2}V_0 I_0 \tag{15}$$

where V_0 and I_0 are voltage and current amplitudes. Equation (15) is the time-average power flow in a lossless line having only an outgoing wave. Hence, the Poynting theorem shows that the power flows through the dielectric of the coaxial line.

16.08. Higher Modes in Coaxial Lines.—If the separation distance between the inner and outer conductors of a coaxial line is of the order of magnitude of a half wavelength or greater, it is possible for higher modes to exist. The field distributions are then of the form given by Eq. (15.05-14). It is necessary to retain the second-kind Bessel function since the field does not extend to the axis, and therefore the singularity of this Bessel function at the axis offers no difficulty. Choosing the cosine variation with respect to ϕ, we write the axial components of intensity as

$$E_z = [C_1 J_n(k\rho) + C_2 N_n(k\rho)] \cos n\phi e^{j\omega t - \Gamma z} \quad TM \text{ modes} \quad (1)$$

$$H_z = [C_1 J_n(k\rho) + C_2 N_n(k\rho)] \cos n\phi e^{j\omega t - \Gamma z} \quad TE \text{ modes} \quad (2)$$

To satisfy the boundary conditions for the TM modes, E_z must be zero at the surfaces of the inner and outer conductors, i.e., at $\rho = a$ and $\rho = b$. This requires that

$$C_1 J_n(ka) + C_2 N_n(ka) = 0$$
$$\qquad\qquad\qquad\qquad\qquad\qquad TM \text{ modes} \qquad (3)$$
$$C_1 J_n(kb) + C_2 N_n(kb) = 0$$

which yields

$$\frac{J_n(ka)}{J_n(kb)} = \frac{N_n(ka)}{N_n(kb)} \qquad TM \text{ modes} \qquad (4)$$

For *TE* modes, the tangential electric intensity is E_ϕ. By inserting H_ρ from Eq. (16.06-7) into (16.06-5), we obtain $E_\phi = [j\omega\mu/(\Gamma^2 - \gamma^2)](\partial H_z/\partial\rho)$.
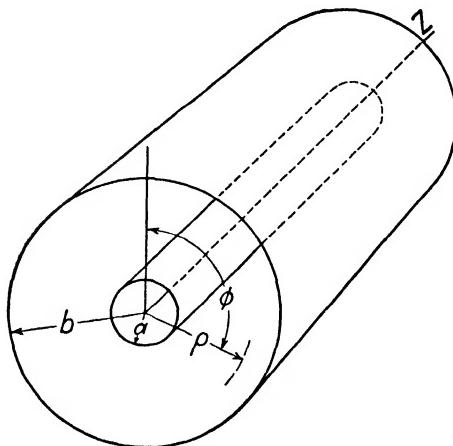
Fig. 8.—Coaxial line.

Consequently, $\partial H_z/\partial\rho$ must be zero at the guide walls in order to make E_ϕ zero. Hence, to satisfy the boundary conditions for *TE* modes, it is necessary that

$$C_1 J'_n(ka) + C_2 N'_n(ka) = 0$$
$$C_1 J'_n(kb) + C_2 N'_n(kb) = 0 \qquad TE \text{ modes} \qquad (5)$$

or

$$\frac{J'_n(ka)}{J'_n(kb)} = \frac{N'_n(ka)}{N'_n(kb)} \qquad TE \text{ modes} \qquad (6)$$

Again we have $k^2 = \Gamma^2 - \gamma^2$, or $\Gamma = \sqrt{k^2 - \omega^2\mu\epsilon}$ for a lossless dielectric, and cutoff occurs when $\omega_0^2\mu\epsilon = k^2$, or

Fig. 9.—Coaxial line.

$$\lambda_0 = \frac{2\pi}{k} \qquad (7)$$
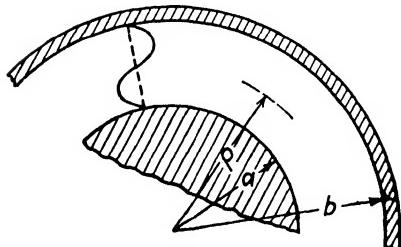
The difficulty arises in the evaluation of k. This may be evaluated by solving Eq. (4) or (6) either graphically or by a cut-and-try process.

For large values of (ka) and (kb) such that $(ka) \gg 1$ and $(kb) \gg 1$, the Bessel functions may be replaced by their asymptotic expressions given in Eqs. (15.06-3 and 4). These indicate a sinusoidal distribution of field intensities in the radial direction. Referring to Fig. 9, we may write the

axial component of electric intensity for *TM* modes in the form

$$E_z = C_1 \sin k(\rho - a) \cos n\phi e^{j\omega t - \Gamma z} \tag{8}$$

where ρ in the denominator of Eq. (15.06-3) is treated as a constant. Since E_z must vanish at the conductor surfaces, we have $k = m\pi/(b - a)$ where m is an integer. Consequently, the cutoff wavelength is

$$\lambda_0 = \frac{2(b - a)}{m} \tag{9}$$

For $m = 1$, the distance $b - a$ must be at least a half wavelength. In general, a given mode in a coaxial line requires a larger radius of outer conductor than is required for the corresponding mode in a circular wave guide.
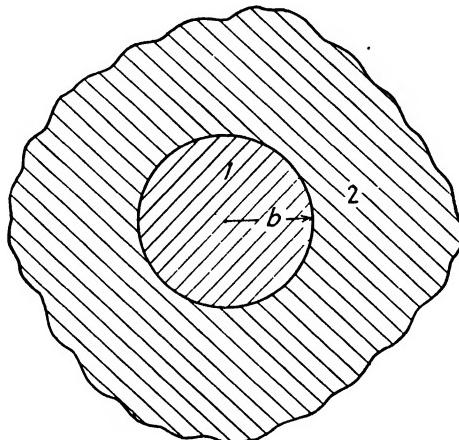
FIG. 10.—Circular wave guide—general case.

16.09. Wave Guides of Circular Cross Section—General Case.—Circular wave guides may take any one of the following forms:

1. A hollow or dielectric filled metallic guide with the electromagnetic field predominantly in the dielectric.

2. A coaxial line with the electromagnetic field in the dielectric between two concentric conducting cylinders.

3. A conducting cylinder in a dielectric medium, with the field in the dielectric outside the cylinder.

4. A dielectric cylinder in a second dielectric medium, with the field in both mediums.

The first two types have been previously considered. We now turn to a more general approach which is applicable to all four types.

Consider an infinitely long circular cylinder, designated medium 1 in Fig. 10, which is immersed in a second medium, designated medium 2. No

restrictions are imposed upon either medium, except that they are homogeneous and isotropic. We assume a wave traveling in the axial direction, with a propagation term $e^{j\omega t - \Gamma z}$. Equation (15.05-14) gives the general form of the intensities in the two mediums. In medium 1, we admit only the first-kind Bessel function to avoid infinite intensity on the axis. In medium 2, however, both the first- and second-kind Bessel functions may exist; hence the field may be represented by the second-kind Hankel function.

The second-kind Hankel function is a linear combination of the first- and second-kind Bessel functions and its asymptotic expression for large values of $(k\rho)$ is given by $H_n^{(2)}(k_2\rho) = \sqrt{(2/\pi k_2\rho)}e^{-j[k_2\rho - (2n+1)\pi/4]}$. The exponential term may be interpreted as representing a wave traveling radially outward from the axis. The axial intensities in the two mediums are

$$E_{z1} = C_1 J_n(k_1\rho)\cos n\phi \qquad E_{z2} = C_3 H_n^{(2)}(k_2\rho)\cos n\phi$$

$$H_{z1} = C_2 J_n(k_1\rho)\cos n\phi \qquad H_{z2} = C_4 H_n^{(2)}(k_2\rho)\cos n\phi \tag{1}$$

where the term $e^{j\omega t - \Gamma z}$ is omitted. The remaining intensity components may be obtained by inserting Eqs. (1) into (16.06-1 to 6). We also have

$$k_1^2 = \Gamma^2 - \gamma_1^2 \qquad k_2^2 = \Gamma^2 - \gamma_2^2 \tag{2}$$

where γ_1 and γ_2 are the intrinsic propagation constants of the mediums. Since the axial propagation constant Γ must be the same in both mediums, it follows that k_1 and k_2 are interrelated. Combining Eqs. (2), we obtain

$$\Gamma^2 = k_1^2 + \gamma_1^2 = k_2^2 + \gamma_2^2$$

$$= \frac{(k_1 b)^2}{b^2} + \gamma_1^2 = \frac{(k_2 b)^2}{b^2} + \gamma_2^2 \tag{3}$$

The tangential components of electric and magnetic intensity are continuous across the boundary unless either medium is a perfect conductor. The field vanishes in a perfectly conducting medium, while in the nonconductor, the tangential electric intensity is zero and the tangential magnetic intensity is equal to the surface-current density. The tangential magnetic intensity satisfies the relationship $\partial H_z/\partial\rho = 0$. Pure transverse-electric and transverse-magnetic modes can exist only when both mediums are lossless dielectrics, or when one medium is a lossless dielectric and the other is a perfect conductor.

Assuming finite conductivity of both mediums, the tangential components of E_z and H_z on either side of the boundary are equated, giving

$$C_1 J_n(k_1 b) = C_3 H_n^{(2)}(k_2 b)$$

$$C_2 J_n(k_1 b) = C_4 H_n^{(2)}(k_2 b) \tag{4}$$

Similar expressions may be obtained by equating E_ϕ and H_ϕ on either side of the boundary. These relationships determine the values of $(k_1 b)$ and $(k_2 b)$. They may be solved either graphically or by a cut-and-try process.

Let us consider the special case in which medium 2 is a perfect conductor. This corresponds to the circular guide considered in Sec. 16.06. For this case, there is no field in medium 2. At the boundary in medium 1, we have $E_z = 0$ for TM modes and $\partial H_z/\partial \rho = 0$ for TE modes. Referring to Eq. (1), it is apparent that these conditions are satisfied by $J_n(k_1 b) = 0$ for TM modes and $J'_n(k_1 b) = 0$ for TE modes. This is in agreement with the relationships previously derived for circular guides. If the guide walls are not perfectly conducting, there will be a field which penetrates into the conductor. It would then be necessary to solve Eqs. (4) for $(k_1 b)$ and $(k_2 b)$ and these values, in turn, are inserted into Eq. (3) to evaluate Γ. If the guide wall is a good conductor, the only appreciable effect upon Γ would be to introduce a small attenuation constant.

Now consider the reverse situation, *i.e.*, medium 1 is a perfectly conducting cylinder and medium 2 is an insulating medium. The field exists only in medium 2. It may be shown [1] that only the symmetrical modes, *i.e.*, those corresponding to $n = 0$, can exist at any appreciable distance from the source. We will consider only the transverse magnetic modes, that is, those modes having $H_z = 0$. To evaluate Γ it would be necessary to set $E_\phi = 0$ and $E_z = 0$. It can be shown that these are satisfied if $H_0^{(2)}(k_2 b)/H_1^{(2)}(k_2 b) = 0$ which, in turn, requires that $k_2 = 0$. Consequently, Eq. (2) gives $\Gamma = \gamma = j\omega\sqrt{\mu_2 \epsilon_2}$ and the propagation constant Γ is equal to the intrinsic propagation constant of the dielectric. The field then propagates with a velocity equal to the velocity of light in the dielectric. The single-conductor transmission line is of little practical value, since the fields extending radially outward will eventually terminate on some conducting medium which makes it effectively a two-conductor transmission line.

The final illustration is that of a cylindrical dielectric rod in a second dielectric medium. In Chap. 14, we found that it is possible for a wave to be totally reflected at a boundary surface between two dielectric mediums. This occurs if the incident wave is in the medium having the higher dielectric constant and the angle of incidence exceeds the angle of total internal reflection. By utilizing this principle, a wave guide may be constructed of a rectangular dielectric slab without any conducting boundaries. If the slab has the proper thickness, the wave will be reflected from wall to wall as it propagates longitudinally down the guide. The same phenomenon also occurs in dielectric guides of circular cross section.

[1] STRATTON, J. A., "Electromagnetic Theory," chap. 9, McGraw-Hill Book Company, Inc., New York, 1941.

For the circular rod, the values of (k_1b) and (k_2b) which determine the value of Γ are again obtained from the boundary equations. The circularly symmetrical case $(n = 0)$ is the simplest, since the Bessel and Hankel functions then reduce to zero-order functions. This is the only case in which the wave may be either a *TE* or *TM* wave, since if $n > 0$, both E_z and H_z are present. We will consider this special case, assuming that both mediums are lossless.

In Eq. (3) we let $\gamma_1^2 = -\omega^2 \mu_1 \epsilon_1$ and $\gamma_2^2 = -\omega^2 \mu_2 \epsilon_2$ to get

$$\Gamma^2 = \frac{(k_1b)^2}{b^2} - \omega^2 \mu_1 \epsilon_1 = \frac{(k_2b)^2}{b^2} - \omega^2 \mu_2 \epsilon_2 \tag{5}$$

from which we obtain

$$\omega = \frac{1}{b} \sqrt{\frac{(k_1b)^2 - (k_2b)^2}{\mu_1 \epsilon_1 - \mu_2 \epsilon_2}} \tag{6}$$

$$\Gamma = \sqrt{\frac{(k_1b)^2}{b^2} - \omega^2 \mu_1 \epsilon_1} = \sqrt{\frac{(k_2b)^2}{b^2} - \omega^2 \mu_2 \epsilon_2} \tag{7}$$

The intensity E_z or H_z in medium 2 has a Hankel function variation which, for large values of $k_2\rho$ may be replaced by the asymptotic representation

$$H_0^{(2)}(k_2\rho) = \sqrt{\frac{2}{\pi(k_2\rho)}} \, e^{-j[k_2\rho - \pi/4]} \tag{8}$$

If the field is to exist predominantly in the dielectric rod, it is necessary that k_2 be imaginary or that $(k_2b)^2$ be negative. The intensities in medium 2 are then attenuated in a direction radially outward from the surface of the cylinder. The borderline case occurs when $(k_2b) = 0$. The corresponding value of ω, from Eq. (6), is

$$\omega = \frac{(k_1b)}{b\sqrt{\mu_1 \epsilon_1 - \mu_2 \epsilon_2}} \tag{9}$$

$$\Gamma = j\beta_2 = j\omega\sqrt{\mu_2 \epsilon_2} \qquad v = \frac{\omega}{\beta_2} = \frac{1}{\sqrt{\mu_2 \epsilon_2}} \tag{10}$$

Hence the propagation constant is that of medium 2 and the wave travels with a velocity characteristic of that of medium 2. The frequency given by Eq. (9) can be considered the cutoff frequency, since it is the frequency for which the radial propagation constant in medium 2 changes from real to imaginary, or from a phase constant to an attenuation constant. Equation (4) shows that (k_1b) is a solution of $J_0(k_1b) = 0$, the lowest value being $(k_1b) = 2.405$. This may be inserted into Eq. (9) to obtain the cutoff frequency.

If $(k_2 b)^2$ is very large and negative, Eq. (6) may be approximated by

$$\omega = \frac{|(k_2 b)|}{b \sqrt{\mu_1 \epsilon_1 - \mu_2 \epsilon_2}} \tag{11}$$

$$\Gamma = j\beta_1 = j\omega \sqrt{\mu_1 \epsilon_1} \qquad v = \frac{\omega}{\beta_1} = \frac{1}{\sqrt{\mu_1 \epsilon_1}} \tag{12}$$

This occurs at high frequencies. The wave propagation takes place predominantly in medium 1, with a velocity and propagation constant determined by medium 1. The field in medium 2 decreases rapidly with distance from the surface and therefore is confined to a thin film at the surface of the dielectric rod.

16.10. Power Transmission Through Wave Guides.—The power transmitted through a wave guide and the power loss in the guide walls may be evaluated by the use of the Poynting theorem. Assume an infinitely long guide, or a guide terminated in such a manner as to avoid reflections at the distant end. The guide has conducting walls and may have either circular or rectangular cross section. In general, there are two pairs of EH vectors which contribute to the longitudinal power flow. Thus, in rectangular guides the two pairs are E_x, H_y and E_y, H_x, which are related by $E_x/H_y = -E_y/H_x = Z_0$. In cylindrical guides, the two pairs are E_ρ, H_ϕ and E_ϕ, H_ρ, these being related by $E_\rho/H_\phi = -E_\phi/H_\rho = Z_0$.

For a lossless dielectric, the time-average power density is obtained by inserting Z_0 for η in Eq. (14.04-3), yielding

$$\mathcal{P}_T = \frac{1}{2Z_0} |E|^2 = \frac{Z_0}{2} |H|^2 \tag{1}$$

where $|E|$ and $|H|$ represent the resultant transverse intensities; thus for rectangular guides $|E|^2 = |E_x|^2 + |E_y|^2$ or $|H|^2 = |H_x|^2 + |H_y|^2$.

Integrating Eq. (1) over the cross section of the guide yields the time-average power flow through the guide, thus

$$P_T = \frac{1}{2Z_0} \int_A |E|^2 \, da = \frac{Z_0}{2} \int_A |H|^2 \, da \tag{2}$$

where the symbol $\int_A da$ denotes integration over the cross section of the guide. For rectangular guides, the first of Eq. (2) becomes

$$P_T = \frac{1}{2Z_0} \int_0^b \int_0^a (|E_x|^2 + |E_y|^2) \, dx \, dy \tag{3}$$

and for circular guides

$$P_T = \frac{1}{2Z_0} \int_0^{2\pi} \int_0^b (|E_\rho|^2 + |E_\phi|^2) \rho \, d\rho \, d\phi \tag{4}$$

Substitution of the intensities for the TE and TM modes as given in Secs 16.04 to 16.06, together with $Z_0 = \eta/\sqrt{1 - (\lambda/\lambda_0)^2}$ for TE modes and $Z_0 = \eta\sqrt{1 - (\lambda/\lambda_0)^2}$ for TM modes, into Eqs. (3) and (4), yields [1]

Rectangular guides

$$P_T = \frac{ab\eta}{8}\left(\frac{\lambda_0}{\lambda}\right)^2\sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2}H_0^2 \qquad TE \text{ modes} \quad (5)$$

Circular guides

$$P_T = \frac{\pi b^2\eta}{4}\left(\frac{\lambda_0}{\lambda}\right)^2\sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2}\left(1 - \frac{n^2}{k^2b^2}\right)H_0^2J_n^2(kb) \quad TE \text{ modes} \quad (6)$$

Rectangular guides

$$P_T = \frac{ab}{8\eta}\left(\frac{\lambda_0}{\lambda}\right)^2\sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2}E_0^2 \qquad TM \text{ modes} \quad (7)$$

Circular guides

$$P_T = \frac{\pi b^2}{4\eta}\left(\frac{\lambda_0}{\lambda}\right)^2\sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2}E_0^2[J_n'(kb)]^2 \quad TM \text{ modes} \quad (8)$$

[1] The time-average power transmitted through the guide may also be expressed in terms of the longitudinal intensities E_z or H_z as given by

$$P_T = \frac{\eta}{2}\left(\frac{\lambda_0}{\lambda}\right)^2\sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2}\int_A|H_z|^2\,da \qquad TE \text{ modes}$$

$$P_T = \frac{1}{2\eta}\left(\frac{\lambda_0}{\lambda}\right)^2\sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2}\int_A|E_z|^2\,da \qquad TM \text{ modes}$$

These integrals are easier to evaluate than those of Eq. (4), particularly for the circular wave guide. To illustrate the method, consider the TM mode in a circular guide, for which Eq. (16.06-16) gives $E_z = E_0J_n(k\rho)\cos n\phi$. Substituting this into the above equation, with $C_1 = (1/2\eta)(\lambda_0/\lambda)^2\sqrt{1 - (\lambda/\lambda_0)^2}$, we obtain

$$P_T = C_1E_0^2\int_0^b\int_0^{2\pi}\rho[J_n(k\rho)]^2\cos^2 n\phi\,d\phi\,d\rho = \pi C_1E_0^2\int_0^b\rho[J_n(k\rho)]^2\,d\rho$$

$$= \pi C_1E_0^2\frac{b^2}{2}\left\{[J_n'(kb)]^2 + \left(1 - \frac{n^2}{k^2b^2}\right)J_n^2(kb)\right\}$$

The last integration is obtained from Eq. (15.10-14). For TM modes, we have $J_n(kb) = 0$, hence

$$P_T = \frac{\pi}{2}C_1E_0^2b^2[J_n'(kb)]^2$$

Replacing C_1 by its expression yields Eq. (8).

For the derivation of power transfer equations similar to those given above, see S. A. Schelkunoff, "Electromagnetic Waves," chap. 10, D. Van Nostrand Company, Inc., New York, 1943.

16.11. Attenuation in Wave Guides.—The attenuation of electromagnetic waves in wave guides may result from one or more of the following causes:

1. An impressed wavelength greater than the cutoff wavelength.
2. Losses in the dielectric.
3. Losses in the guide walls.

If the impressed wavelength is greater than cutoff, the intensities will be attenuated even though the guide is lossless. This type of attenuation is due to internal reflection at the entrance of the guide. The effect is similar to the attenuation experienced in lossless filters when operating in the attenuation band. For this type of attenuation, the propagation constant, as given by Eq. (16.03-14), is real, and we set $\Gamma = \alpha_1$, as indicated in Eq. (16.03-21). If the impressed wavelength is much greater than cutoff, α_1 approaches the limiting value of

$$\alpha_1 = \frac{2\pi}{\lambda_0} \tag{1}$$

Wave guides with dimensions much smaller than cutoff are often used as attenuators. The input and output circuits may consist of coaxial lines which are coupled to the guide by means of either probe or loop antennas. The attenuation can be varied by means of a sliding piston which varies the length of the guide.

16.12. Attenuation Due to Dielectric Losses.—Let us now consider the attenuation of a wave resulting from dielectric losses. For a plane wave traveling in unbounded dielectric, the propagation constant, as given by Eq. (13.06-5), is $\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)}$. This may be written

$$\gamma = \sqrt{-\omega^2\mu\epsilon\left(1 - j\frac{\sigma}{\omega\epsilon}\right)} = j\omega\sqrt{\mu\epsilon}\left(1 - j\frac{\sigma}{\omega\epsilon}\right)^{\frac{1}{2}} \tag{1}$$

The second form of Eq. (1) may be expanded by the binomial theorem. For a low-loss dielectric, *i.e.*, $\sigma/\omega\epsilon \ll 1$, only the first two terms of the series are significant. With the substitution $\omega^2\mu\epsilon = (2\pi/\lambda)^2$, these terms may be written

$$\gamma = \frac{\pi}{\lambda}\left(\frac{\sigma}{\omega\epsilon}\right) + j\omega\sqrt{\mu\epsilon} \tag{2}$$

The phase constant is the familiar $\beta = \omega\sqrt{\mu\epsilon}$. The attenuation constant may be expressed as

$$\alpha = \frac{\pi}{\lambda}\left(\frac{\sigma}{\omega\epsilon}\right) = \frac{\eta\sigma}{2} \tag{3}$$

where η is the intrinsic impedance of the dielectric. The approximate power factor of a low-loss dielectric is $P.F. = \sigma/\omega\epsilon$. Hence, if the power

factor of the dielectric is known, the attenuation constant may be readily computed from the first form of Eq. (3).

A similar approach may be used to evaluate the attenuation in wave guides resulting from dielectric losses. The propagation constant in the guide is

$$\Gamma = \sqrt{\gamma^2 + k^2} \qquad (4)$$

Inserting γ from Eq. (1) with $\omega^2 \mu \epsilon = (2\pi/\lambda)^2$ and $k^2 = (2\pi/\lambda_0)^2$, we have

$$\Gamma = \frac{2\pi}{\lambda} \left\{ \left[\left(\frac{\lambda}{\lambda_0}\right)^2 - 1 \right] + j\frac{\sigma}{\omega\epsilon} \right\}^{\frac{1}{2}} \qquad (5)$$

Expanding Eq. (5) by the binomial series and retaining the first two terms, we obtain

$$\Gamma = \frac{\pi}{\lambda\sqrt{1 - (\lambda/\lambda_0)^2}} \frac{\sigma}{\omega\epsilon} + j\frac{2\pi}{\lambda} \sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2} \qquad (6)$$

Hence the phase constant in the guide is $\beta_1 = (2\pi/\lambda)\sqrt{1 - (\lambda/\lambda_0)^2}$. The attenuation constant may be expressed as

$$\alpha_1 = \frac{\eta\sigma}{2\sqrt{1 - (\lambda/\lambda_0)^2}} \qquad (7)$$

As λ/λ_0 approaches zero, the attenuation constant in the guide approaches that for unbounded dielectric given by Eq. (3). The attenuation becomes very large as the wavelength of the impressed signal approaches cutoff.

16.13. Attenuation Resulting from Losses in the Guide Walls.—A useful relationship may be obtained which relates the attenuation constant of a guide to the power transmitted through the guide and the power loss. Since this expression will be helpful in evaluating the attenuation constant resulting from losses in the guide walls, let us consider it for a moment.

The time-average power transmitted through the guide is expressed by Eq. (16.10-2). The magnitude of the resultant magnetic intensity may be written $|H| = |H_0|e^{-\alpha_1 z}$, where H_0 is the intensity at $z = 0$. Upon inserting this into Eq. (16.10-2), we obtain

$$P_T = \frac{Z_0}{2} \int_A |H_0|^2 e^{-2\alpha_1 z}\, da \qquad (1)$$

The power loss per unit length of guide is the space rate of decrease of power flow, or $P_L = -\partial P_T/\partial z$. Inserting P_T from Eq. (1) into this relationship, we have

$$P_L = \alpha_1 Z_0 \int_A |H_0|^2 e^{-2\alpha_1 z}\, da \qquad (2)$$

If we now divide Eq. (2) by (1) and solve for α_1, we obtain the desired relationship,

$$\alpha_1 = \frac{P_L}{2P_T} \tag{3}$$

The attenuation constant is therefore equal to the power loss per unit length of guide, divided by twice the transmitted power. This expression is analogous to Eq. (8.03-9) which was derived for the transmission line.

The attenuation constant resulting from the power loss in the guide walls is found by obtaining expressions for P_L and P_T and inserting these into Eq. (3). We have previously derived expressions for the transmitted power. To evaluate the power loss in the guide walls, we first write the equations for the electric and magnetic intensity as though the guide were lossless. Since the walls of the guide are assumed to be perfectly conducting, the tangential component of electric intensity at the guide walls is zero. However, there will be a tangential component of magnetic intensity which is terminated by a current flowing on the surface of the conductor. If we now assume that the guide walls are imperfectly conducting, the tangential component of magnetic intensity will be substantially the same as if the walls were perfectly conducting. We may therefore use Eq. (14.04-4) to evaluate the power density in the imperfectly conducting walls. (The power loss per unit length of guide is obtained by integrating the power density over the surface of the conductor corresponding to unit length of guide; thus, from Eqs. (14.04-4) and (14.07-4), we obtain

$$P_L = \frac{\mathcal{R}_s}{2} \int_s |H_t|^2 \, ds \tag{4}$$

where H_t is the tangential component of magnetic intensity at the wall of the guide and $\mathcal{R}_s = \sqrt{\omega\mu_2/2\sigma_2}$ is the skin-effect resistance of the conductor.

Inserting P_T from Eq. (16.10-2) and P_L above into Eq. (3), we have

$$\alpha_1 = \frac{\mathcal{R}_s \int_s |H_t|^2 \, ds}{2Z_0 \int_A |H|^2 \, da} \tag{5}$$

The integral in the numerator is evaluated over the surface of the guide walls for unit length of guide, whereas that in the denominator is over the cross section of the guide.

Let us now apply the foregoing method to the determination of α_1 for *TM* modes in rectangular guides. The tangential magnetic intensities at

the walls of the guide, as obtained from Eq. (16.05-9 and 10), are

$$| H_{xt} | = \frac{\beta_1}{Z_0} \left(\frac{\lambda_0}{2\pi} \right)^2 k_y E_0 \sin k_x x \tag{6}$$

$$| H_{yt} | = \frac{\beta_1}{Z_0} \left(\frac{\lambda_0}{2\pi} \right)^2 k_x E_0 \sin k_y y \tag{7}$$

These result in a longitudinal current in the guide walls, causing the power loss. Inserting the tangential intensities into Eq. (4), we obtain the power loss

$$P_L = \mathcal{R}_s \left[\frac{\beta_1}{Z_0} \left(\frac{\lambda_0}{2\pi} \right)^2 E_0 \right]^2 \left[k_y^2 \int^a \sin^2 k_x x \, dx + k_x^2 \int_0^b \sin^2 k_y y \, dy \right]$$

$$= \frac{\mathcal{R}_s \lambda_0^4 E_0^2}{8 \lambda^2 \eta^2} \left(\frac{n^2 a}{b^2} + \frac{m^2 b}{a^2} \right) \tag{8}$$

To obtain the attenuation constant, insert P_L from Eq. (8) and P_T from Eq. (16.10-7) into (3), yielding

$$\alpha_1 = \frac{\mathcal{R}_s \lambda_0^2}{2 \eta a b \sqrt{1 - (\lambda/\lambda_0)^2}} \left(\frac{n^2 a}{b^2} + \frac{m^2 b}{a^2} \right) \tag{9}$$

In the case of *TE* modes, the longitudinal magnetic intensity at the guide walls contributes to the power loss resulting from a transverse current flowing in the guide walls. As an illustration, consider the attenuation in a circular guide, $TE_{n,m}$ mode. From Eqs. (16.06-15) the tangential magnetic intensities are

$$| H_{\phi t} | = \frac{\beta_1 n}{k^2 b} H_0 J_n(kb) \sin n\phi \tag{10}$$

$$| H_{zt} | = H_0 J_n(kb) \cos n\phi \tag{11}$$

Inserting these into Eq. (16.13-4) yields the power loss

$$P_L = \frac{\mathcal{R}_s H_0^2}{2} J_n^2(kb) \left[\left(\frac{\beta_1 n}{k^2 b} \right)^2 \int_0^{2\pi} b \sin^2 n\phi \, d\phi + \int_0^{2\pi} b \cos^2 n\phi \, d\phi \right]$$

$$= \frac{\pi b \mathcal{R}_s}{2} H_0^2 J_n^2(kb) \left[\left(\frac{\beta_1 n}{k^2 b} \right)^2 + 1 \right] \tag{12}$$

Substitution of P_L from Eq. (12) and P_T from Eq. (16.10-6) into Eq. (3) yields

$$\alpha_1 = \frac{\mathcal{R}_s}{\eta b \sqrt{1 - (\lambda/\lambda_0)^2}} \left[\left(\frac{\lambda}{\lambda_0} \right)^2 + \frac{n^2}{(kb)^2 - n^2} \right] \tag{13}$$

The attenuation constants of rectangular and circular guides for *TE* and *TM* modes are given in Table 6.   Figure 11 shows the variation of attenuation constant as a function of frequency for a circular guide 5 inches in diameter. It is interesting to note that the attenuation of the $TE_{0,m}$ mode in circular guides decreases indefinitely as the impressed wavelength decreases.   This is evident in Eq. (13), since $\sqrt{1 - (\lambda/\lambda_0)^2}$ approaches unity,
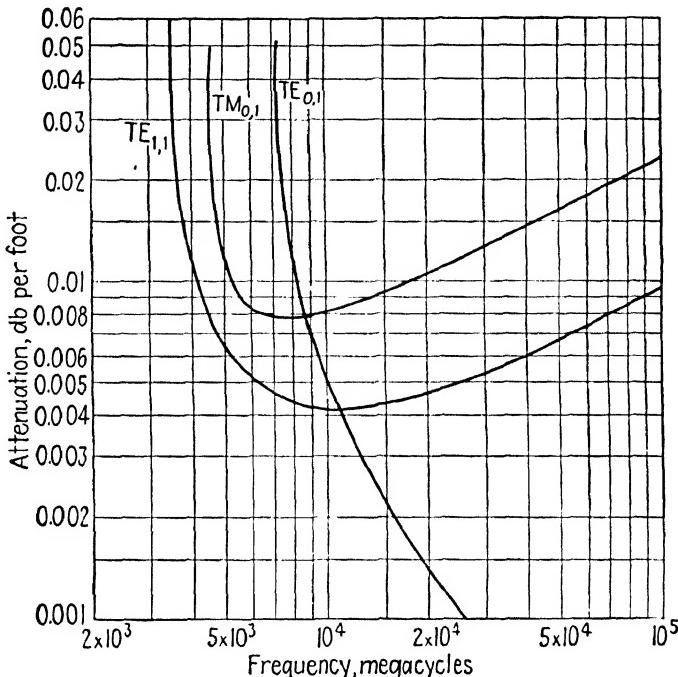
Fig. 11.—Attenuation as a function of frequency in a circular guide 5 inches in diameter.

while (λ/λ_0) decreases approaching zero.   However, L. J. Chu has shown that this anomalous attenuation characteristic is lost if the guide is slightly deformed.

All modes in rectangular or circular guides have an impressed wavelength for which the attenuation is a minimum.   For *TM* modes in either rectangular or circular guides, the optimum wavelength is given by

$$\left(\frac{\lambda_0}{\lambda}\right)_{\text{opt.}} = \left(\frac{f}{f_0}\right)_{\text{opt.}} = \sqrt{3} \tag{14}$$

For *TE* modes, we have

$$\left(\frac{\lambda_0}{\lambda}\right)_{\text{opt.}} = \left(\frac{f}{f_0}\right)_{\text{opt.}} = \sqrt{\frac{3}{2}(1+p) + \sqrt{\frac{9}{4}(1+p)^2 - p}} \tag{15}$$

where

$p = 2a/b$ in rectangular guides $TE_{0,n}$ mode

$p = a/b[1 + (m/n)^2(b/a)^3]/[1 + (m/n)^2(b/a)]$ in   rectangular   guides
$TE_{m,n}$ mode (if $m \neq 0$ and $n \neq 0$)

$p = [(kb)^2 - n^2]/n^2$ in circular guides $TE_{n,m}$ modes $n \neq 0$

A plot of $(\lambda_0/\lambda)_{\text{opt.}}$ as function of p is shown in Fig. 12. As an example, the $TE_{0,n}$ mode in a rectangular guide with a ratio $a/b = 0.8$ has a value of
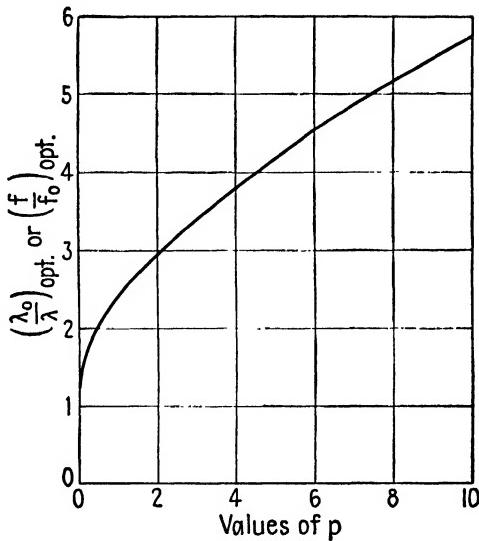
Fig. 12.—Plot of $\left(\dfrac{\lambda_0}{\lambda}\right)_{\text{opt.}}$ as a function of p for rectangular and circular guides (except $TE_{0,m}$ modes in circular guides).

$p = 2a/b = 1.6$. Referring to Fig. 12, we find that minimum attenuation occurs when $(\lambda_0/\lambda)_{\text{opt.}} = (f_0/f)_{\text{opt.}} = 2.75$.

For a given guide perimeter, the dominant mode in circular guides ($TE_{1,1}$ mode) has the lowest attenuation. The next lowest is the $TE_{0,1}$ mode in rectangular guides having a ratio $b/a = 1.18$, the attenuation constant being about 85 per cent greater than the $TE_{1,1}$ mode in circular guides. We might have anticipated lower attenuation in circular guides, since the transmitted power flows through the cross section of the guide, whereas the power loss occurs in the guide walls. Consequently, as a rough criterion, the ratio of cross-sectional area to perimeter should be a maximum for minimum attenuation, a condition which is satisfied by circular guides.

It is interesting to compare the attenuation in circular wave guides with the attenuation of the principal mode in coaxial lines. Assuming the optimum ratio of $b/a = 3.6$ for the line, and the dominant mode in the circular guide, the attenuation ratio, in terms of the impressed wavelength λ and the cutoff wavelength λ_0 in the guide, is

$$\frac{\alpha_{\text{cable}}}{\alpha_{\text{guide}}} = \frac{1.8\sqrt{1 - (\lambda/\lambda_0)^2}}{0.418 + (\lambda/\lambda_0)^2} \tag{16}$$

The attenuation constants are equal when $\lambda/\lambda_0 = 0.88$. For a larger wavelength ratio, the coaxial cable offers lower attenuation, whereas for a lower ratio the guide has less attenuation. As $\lambda/\lambda_0 \to 0$, the *attenuation ratio* given in Eq. (16) approaches the value 4.3. In the wavelength range where guides have approximately the same dimensions as coaxial cables, the guides are preferable, since they offer the advantages of less copper and simpler construction.

Some wave guide modes are relatively more stable than others with slight deformations of the guide walls. L. J. Chu [1] has shown that, for guides of elliptical cross section, (1) all circular modes, *i.e.*, modes in which $n = 0$, are stable under slight changes of cross section along an axis of symmetry, (2) the $TE_{0,1}$ and $TM_{0,1}$ modes are stable under slight changes in cross section, and (3) with the exception of (1) and (2) above, all modes are unstable under cross-section deformations. Also, as previously stated, the attenuation characteristic of $TE_{0,n}$ mode changes with deformations.

[1] Chu, L. J., Electromagnetic Waves in Elliptic Hollow Pipes of Metal, *Jour. Applied Phys.*, vol. 9, pp. 583–591; September, 1938.

TABLE 6.—SUMMARY OF WAVE GUIDE FORMULAS

Cutoff wavelength—rectangular guides
$$\lambda_0 = \frac{2}{\sqrt{(m/a)^2 + (n/b)^2}}$$

Cutoff wavelength—circular guides
$$\lambda_0 = \frac{2\pi b}{(kb)}$$

Where

$$J_n(kb) = 0 \qquad TM \text{ modes}$$

$$J_n'(kb) = 0 \qquad TE \text{ modes}$$

Phase constant in the guide (longitudinal)
$$\beta_1 = \frac{2\pi}{\lambda} \sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2}$$

Wavelength in the guide (longitudinal)
$$\lambda_p = \frac{2\pi}{\beta_1} = \frac{\lambda}{\sqrt{1 - (\lambda/\lambda_0)^2}}$$

Phase velocity (longitudinal)
$$v_p = f\lambda_p = \frac{v_c}{\sqrt{1 - (\lambda/\lambda_0)^2}}$$

Group velocity (longitudinal)
$$v_g = \frac{1}{d\beta_1/d\omega} = v_c \sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2}$$

Velocity relationships
$$v_c = \sqrt{v_g v_p}$$

Characteristic wave impedance—TE modes
$$Z_0 = \frac{\eta}{\sqrt{1 - (\lambda/\lambda_0)^2}}$$

Characteristic wave impedance—TM modes
$$Z_0 = \eta \sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2}$$

Attenuation at wavelengths longer than cutoff (nepers per meter)
$$\alpha_1 = \frac{2\pi}{\lambda} \sqrt{\left(\frac{\lambda}{\lambda_0}\right)^2 - 1}$$

Attenuation due to dielectric loss (nepers per meter)
$$\alpha_1 = \frac{\eta\sigma}{2\sqrt{1 - (\lambda/\lambda_0)^2}}$$

Power transmitted

Rectangular Guides

$$P_T = \frac{ab\eta H_0^2}{8} \left(\frac{\lambda_0}{\lambda}\right)^2 \sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2} \qquad TE \text{ modes}$$

$$P_T = \frac{ab E_0^2}{8\eta} \left(\frac{\lambda_0}{\lambda}\right)^2 \sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2} \qquad TM \text{ modes}$$

Circular Guides

$$P_T = \frac{\pi b^2 \eta H_0^2}{4} \left(\frac{\lambda_0}{\lambda}\right)^2 \left(1 - \frac{n^2}{k^2 b^2}\right) J_n^2(kb) \sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2} \qquad TE \text{ modes}$$

$$P_T = \frac{\pi b^2 E_0^2}{4\eta} \left(\frac{\lambda_0}{\lambda}\right)^2 [J_n'(kb)]^2 \sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2} \qquad TM \text{ modes}$$

Table 6.—Summary of Wave Guide Formulas (*Continued*)

Attenuation in hollow guides due to losses in the guide walls (decibels per meter)

Rectangular Guides

$$\frac{0.79}{a\nu}\sqrt{\frac{\mu_r}{\sigma_2\lambda}}\left[1 + \frac{2a}{b}\left(\frac{\lambda}{\lambda_0}\right)^2\right] \qquad TE_{0,n}\text{ modes}$$

$$\frac{1.59}{b\nu}\sqrt{\frac{\mu_r}{\sigma_2\lambda}}\left\{\frac{\dfrac{b}{a}\left[\dfrac{b}{a}\left(\dfrac{m}{n}\right)^2 + 1\right]\nu^2}{\left(\dfrac{b}{a}\right)^2\left(\dfrac{m}{n}\right)^2 + 1} + \left(1 + \frac{b}{a}\right)\left(\frac{\lambda}{\lambda_0}\right)^2\right\} \qquad TE_{m,n}\text{ modes}$$

$$\frac{0.79\lambda_0^2}{2b^3\nu}\sqrt{\frac{\mu_r}{\sigma_2\lambda}}\left[\left(\frac{b}{a}\right)^3 m^2 + n^2\right] \qquad TM_{m,n}\text{ modes}$$

Circular Guides

$$\frac{0.79}{b\nu}\sqrt{\frac{\mu_r}{\sigma_2\lambda}}\left[\left(\frac{\lambda}{\lambda_0}\right)^2 + \frac{n^2}{(kb)^2 - n^2}\right] \qquad TE_{n,m}\text{ modes}$$

$$\frac{0.79}{b\nu}\sqrt{\frac{\mu_r}{\sigma_2\lambda}} \qquad TM_{n,m}\text{ modes}$$

Coaxial Line

$$\frac{0.79}{b}\sqrt{\frac{\mu_r}{\sigma_2\lambda}}\left[\frac{1 + \dfrac{b}{a}}{2\ln\dfrac{b}{a}}\right]$$

where

$$\nu = \sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2}$$

$$\eta = \sqrt{\frac{\mu}{\epsilon}}$$

μ_r = relative permeability (unity for non-magnetic materials)

PROBLEMS

1. A hollow rectangular guide has dimensions $a = 4$ cm, and $b = 6$ cm. The frequency of the impressed signal is 3,000 megacycles per sec. Compute the following for the $TE_{0,1}$, $TE_{1,1}$, and $TE_{2,2}$ modes:
   (*a*) Cutoff wavelength.
   (*b*) Wavelength parallel to the guide walls for modes in the pass band.
   (*c*) Phase velocity and group velocity for modes in the pass band.
   (*d*) Characteristic wave impedance for modes in the pass band.
   (*e*) Attenuation constant for modes in the attenuation band.

2. Show that the group velocity in a lossless wave guide is given by

$$v_g = v_c \sqrt{1 - \left(\frac{\lambda}{\lambda_0}\right)^2}$$

by inserting β from Eq. (16.03-11) into Eq. (14.14-7).

3. Starting with the axial magnetic intensity, as given by Eq. (16.04-8), derive expressions for the remaining electric and magnetic intensities in a rectangular guide for the *TE* modes.

4. Show that if both E_z and H_z are zero in a rectangular guide, all of the other intensities are zero and hence the *TEM* mode cannot exist in a guide. What happens to the axial component of electric or magnetic intensity as the dimensions of the guide are increased indefinitely?

5. A circular wave guide is to be operated at a frequency of 5,000 megacycles per sec and is to have dimensions such that $\lambda/\lambda_0 = 0.9$ for the dominant mode. Evaluate the following:

(a) The diameter of the guide.

(b) The values of λ_p, v_p, v_g, and Z_0.

(c) The attenuation in decibels per meter for the next higher mode in the guide.

6. Starting with the axial electric intensity, as given by Eq. (16.06-12), derive the expressions for the remaining electric and magnetic intensities in a circular guide for the *TM* modes.

7. The attenuation of a hollow wave guide is to be compared with that of a dielectric-filled guide at a frequency of 3,000 megacycles per sec. Both guides are silver plated and have rectangular shape with a height equal to two-thirds of the width. The guides are operated in the $TE_{0,1}$ mode and are each designed such that $\lambda/\lambda_0 = 0.8$. The dielectric-filled guide has a dielectric having the properties $\epsilon_r = 2.5$ and $\sigma/\omega\epsilon = 0.0005$.

(a) Specify the dimensions of the two guides.

(b) Compute the attenuation constants and the attenuation in decibels per meter.

8. Compute the cutoff wavelength and corresponding frequencies for the first three higher modes in a coaxial line having dimensions $a = 3$ cm, $b = 4$ cm, considering only those modes of the type expressed by Eq. (16.08-8).

9. A rectangular wave guide is designed to have a ratio $\lambda/\lambda_0 = 0.8$ at a frequency of 5,000 megacycles per sec in the $TE_{0,1}$ mode. The guide has a height to width ratio of 0.5. The time-average power flow through the guide is 1 kw. Compute the maximum values of electric and magnetic intensities in the guide and indicate where these occur in the guide. Evaluate the currents in each of the side walls and indicate the directions in which they flow.

10. Derive Eq. (16.10-6) for the power flow in a circular guide *TE* mode.

11. Show that $\nabla \cdot \bar{E} = 0$ and $\nabla \cdot \bar{H} = 0$ for *TE* and *TM* modes in rectangular guides.

12. An infinitely long dielectric slab of thickness b is immersed in a second dielectric medium. Both mediums are assumed to be lossless. Show that the cut-off wavelength for the slab as a wave guide is

$$\lambda_0 = 2b\sqrt{1 - (\epsilon_2/\epsilon_1)}$$

Hint: Assume that at cutoff the angle of incidence in the slab is equal to the angle of total internal reflection, given by Eq. (14.12-3) and that the wavelength normal to the guide is $2b$.

CHAPTER 17

IMPEDANCE DISCONTINUITIES IN GUIDES—RESONATORS

The resultant wave traveling down an infinitely long guide, which has a uniform characteristic wave impedance, may be regarded as an outgoing wave. Since this outgoing wave sees no discontinuity in wave impedance, there is no tendency to set up a reflected wave. It is also possible to terminate a uniform guide of finite length in such a manner that the outgoing wave sees no impedance discontinuity, thereby eliminating reflections. These two cases are analogous to the transmission line which is infinitely long and the line which is terminated in its characteristic impedance, respectively.

17.01. Effect of Impedance Discontinuities in Guides.—The effect of impedance discontinuities in wave guides is to produce reflections which originate at the points of impedance discontinuity. The outgoing and reflected waves, traveling in opposite directions, produce standing waves of electric and magnetic intensity in the guide. These standing waves are analogous to the standing waves of voltage and current on a transmission line resulting from an impedance discontinuity on the line.

The standing-wave ratio in the guide (ratio of the maximum to minimum value of electric intensity) may be measured by means of a traveling detector such as that shown in Fig. 6, Chap. 18. The methods described in Sec. 10.05 may be used to evaluate wave impedances in terms of the standing-wave ratio. A standing-wave ratio of unity indicates that there is no reflected wave; hence the energy of the outgoing wave must be totally absorbed in the load. A standing-wave ratio appreciably greater than unity indicates that the guide contains an abrupt impedance discontinuity. From a practical point of view, it is often possible, by careful adjustment, to obtain standing-wave ratios of the order of 1.01.

There are a number of ways of terminating wave guides so as to avoid reflected waves. For example, a wave guide may be terminated by a metallic horn such as that shown in Fig. 7, Chap. 21. The electromagnetic horn provides a gradual transformation of impedance from the characteristic wave impedance of the guide to the intrinsic impedance of free space. If the electromagnetic horn is many wavelengths long, there will be virtually no reflected wave and all of the energy of the outgoing wave will be radiated into space.

Another method of terminating a guide so as to utilize the energy of the outgoing wave is to place a small pickup antenna in the guide and connect

this, through a coaxial line, to an external load. Any one of the impedance-matching methods described in Chap. 10 may be used to match the impedance of the load to the impedance of the guide, thereby assuring maximum power transfer from the guide to the load. To prevent loss of energy due to radiation, the guide may be closed off by a conducting wall placed across the end of the guide just beyond the pickup antenna. Although a pickup antenna of this type does not provide a uniform termination across the guide, experience has shown that it is possible to obtain a standing-wave ratio approaching unity by this means.

Consider a wave guide which is excited in such a manner that the outgoing wave travels down the guide in the dominant mode. If the guide contains an impedance discontinuity, the reflected wave can propagate either in the dominant mode, or in a superposition of modes including the dominant mode and higher-order modes. Whether or not the higher-order modes will appear in the reflected wave depends upon the nature of the impedance discontinuity and the dimensions of the guide.

If the impedance discontinuity is uniform across the guide, then there is little likelihood that higher-order modes will appear in the reflected wave. Such a uniform impedance discontinuity might arise from the use of two different dielectric mediums in the guide, with a plane interface between the two mediums which is transverse to the guide, as shown in Fig. 1.

If the impedance discontinuity is nonuniform, such as that provided by an aperture or an antenna in the guide, then higher-order modes will appear in the vicinity of the impedance discontinuity. If the dimensions of the guide are such as to pass the dominant modes and attenuate all higher-order modes, then the higher-order modes cannot be present in the reflected or transmitted waves at any appreciable distance from the impedance discontinuity. However, if the dimensions of the guide are such as to pass some of the higher-order modes, then the reflected wave will in all probability contain the dominant mode as well as some of the higher-order modes; hence there will be energy transfer from the dominant mode to the higher-order modes.

17.02. Wave Guide with Two Different Dielectric Mediums.—Consider an infinitely long guide which contains two different dielectric mediums with a plane interface normal to the axis of the guide as shown in Fig. 1. It will be assumed that the outgoing, reflected, and transmitted waves all propagate in the dominant mode. The outgoing and transmitted waves are assumed to travel in the $-z$ direction and therefore contain the propagation term $e^{\Gamma z}$, while the reflected wave travels in the $+z$ direction and contains the propagation term $e^{-\Gamma z}$.

For either a TE or TM mode in a rectangular guide, there will be two pairs of transverse intensity components which contribute to the longitudinal component of Poynting's vector. These are E_x, H_y and E_y, H_x.

The intensities of the outgoing and reflected waves for the first pair in the medium 1 may be represented by

$$E_x = (Ae^{\Gamma_1 z} + Be^{-\Gamma_1 z})f(x, y) \tag{1}$$

$$H_y = \left(\frac{A}{Z_{01}} e^{\Gamma_1 z} - \frac{B}{Z_{01}} e^{-\Gamma_1 z}\right)f(x, y) \tag{2}$$

The time function $e^{j\omega t}$ has been omitted in the above equations for brevity, and the transverse spatial distribution function $f(x, y)$ may be obtained from Eqs. (16.04-21) for *TE* modes or (16.05-7) for *TM* modes.

Outgoing wave
Reflected wave
Γ_1, Z_{01}
Γ_2, Z_{02}
Transmitted wave
z
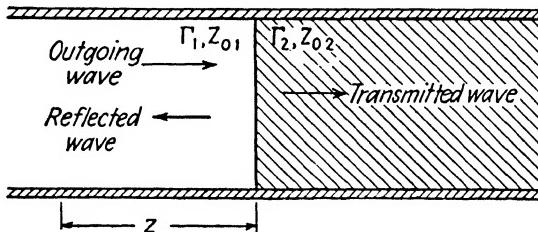
FIG. 1.—Wave guide with two different dielectric mediums.

The A and B coefficients in Eqs. (1) and (2) may be evaluated in terms of the surface intensities by the methods used for transmission lines in Sec. 8.02 or for plane-wave reflections in Sec. 14.05. The wave impedance terminating medium 1 is the characteristic wave impedance Z_{02} of the second medium. Letting E_{xR} be the value of E_x at the interface between the two dielectric mediums, the A and B coefficients are given by the expressions.

$$A = \frac{E_{xR}}{2}\left(1 + \frac{Z_{01}}{Z_{02}}\right) \qquad B = \frac{E_{xR}}{2}\left(1 - \frac{Z_{01}}{Z_{02}}\right) \tag{3}$$

Upon inserting these into Eqs. (1) and (2), we obtain

$$E_x = \left[\frac{E_{xR}}{2}\left(1 + \frac{Z_{01}}{Z_{02}}\right)e^{\Gamma_1 z} + \frac{E_{xR}}{2}\left(1 - \frac{Z_{01}}{Z_{02}}\right)e^{-\Gamma_1 z}\right]f(x, y) \tag{4}$$

$$H_y = \left[\frac{E_{xR}}{2Z_{01}}\left(1 + \frac{Z_{01}}{Z_{02}}\right)e^{\Gamma_1 z} - \frac{E_{xR}}{2Z_{01}}\left(1 - \frac{Z_{01}}{Z_{02}}\right)e^{-\Gamma_1 z}\right]f(x, y) \tag{5}$$

Equations (4) and (5) are analogous to Eqs. (14.05-2 and 3) for the reflection of uniform plane waves. In order to satisfy the boundary conditions, the tangential electric and magnetic intensities must be equal on either side of the boundary surface between the two dielectric mediums. Since E_{xR} is the value of E_x at the boundary surface in medium 1, it must likewise be the value of E_x at the boundary surface in medium 2. The intensi-

ties of the outgoing wave in medium 2 may therefore be written

$$E_x''' = E_{xR}e^{\Gamma_2 z}f(x, y) \tag{6}$$

$$H_y''' = \frac{E_{xR}}{Z_{02}} e^{\Gamma_2 z}f(x, y) \tag{7}$$

Expressions similar to those given by Eqs. (1) to (7) may be written for the E_y, H_x components of intensity. The reflection and transmission coefficients are given by Eqs. (14.10-17 and 18) which, for this case, become

$$r_R = \frac{Z_{02} - Z_{01}}{Z_{02} + Z_{01}} \tag{8}$$

$$r_T = \frac{2Z_{02}}{Z_{02} + Z_{01}} \tag{9}$$

If both mediums are lossless dielectrics, the characteristic wave impedances are given by Eqs. (16.04-16) for *TE* modes and Eq. (16.05-4) for *TM* modes.

When Eqs. (4) and (5) are expressed in hyperbolic form, they become

$$E_x = E_{xR} \left(\cosh \Gamma_1 z + \frac{Z_{01}}{Z_{02}} \sinh \Gamma_1 z \right) f(x, y) \tag{10}$$

$$H_y = \frac{E_{xR}}{Z_{02}} \left(\cosh \Gamma_1 z + \frac{Z_{02}}{Z_{01}} \sinh \Gamma_1 z \right) f(x, y) \tag{11}$$

and the wave impedance for the transverse components of intensity may be written

$$Z = \frac{E_x}{H_y} = Z_{01} \left(\frac{Z_{02} + Z_{01} \tanh \Gamma_1 z}{Z_{01} + Z_{02} \tanh \Gamma_1 z} \right) \tag{12}$$

17.03. Wave Guide with a Perfectly Conducting End Wall.—The relationships derived in the preceding section apply equally well if medium 2 is a dielectric or a conducting medium. As a special case, consider a rectangular guide which has a lossless dielectric and perfectly conducting side walls. The guide is also assumed to be terminated by a perfectly conducting end wall, which we designate as medium 2. We then have $\Gamma_1 = j\beta_1$, $Z_{02} = 0$, and $E_{xR} = 0$. Making these substitutions in Eqs. (17.02-10) to (12), and using $H_{yR} = E_{xR}/Z_{02}$ to eliminate the indeterminate, we obtain

$$E_x = jH_{yR}Z_{01} \sin \beta_1 z \, f(x, y) \tag{1}$$

$$H_y = H_{yR} \cos \beta_1 z \, f(x, y) \tag{2}$$

$$Z = \frac{E_x}{H_y} = jZ_{01} \tan \beta_1 z \tag{3}$$

Equations (1) and (2) are the equations for standing waves of electric and magnetic intensity in the guide. The standing waves are shown in Fig. 2. These equations and the standing-wave patterns are similar to those for the voltage and current on a short-circuited lossless transmission line.
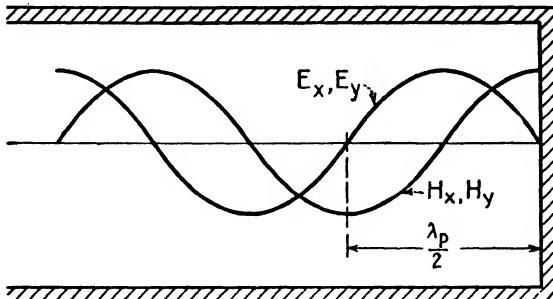
Fig. 2.—Standing waves in a guide.

The wave impedance given by Eq. (3) is reactive, indicating that the electric and magnetic intensities are in time quadrature and that the time-average power is zero. The variation of reactance as a function of distance z from the end wall (or as a function of $\beta_1 z$) is similar to that shown in Fig. 5, Chap. 8.

17.04. Impedance Matching Using a Dielectric Slab.—In transmission-line theory it was shown that a quarter-wavelength section of line may be used to obtain an impedance match if the generator and load impedances are pure resistances. If a wave guide contains two different dielectric
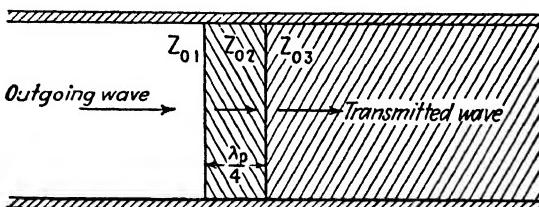
Fig. 3.—Impedance matching by use of a quarter-wavelength dielectric slab.

mediums, it is possible to obtain an impedance match and thereby avoid reflections by interposing a dielectric slab between the two original dielectric mediums, as shown in Fig. 3. The dielectric slab must be a quarter-wavelength thick (as measured in the guide) and have a characteristic wave impedance given by

$$Z_{02} = \sqrt{Z_{01} Z_{03}} \tag{1}$$

If the impedances are matched, there will be an outgoing wave in medium 1, but no reflected wave. All of the energy of the outgoing wave will then be transmitted to medium 3.

17.05. Apertures in Wave Guides.—If a wave guide contains a non-uniform discontinuity, such as the step discontinuity of Fig. 4 or the apertures of Fig. 5, new modes may appear in the reflected and transmitted waves which are not present in the outgoing wave.

Assume that the outgoing wave in Fig. 4 propagates in the dominant mode and that the dimensions of the guide are such as to attenuate all higher-order modes. Field distributions corresponding to higher-order

(a)

(b)
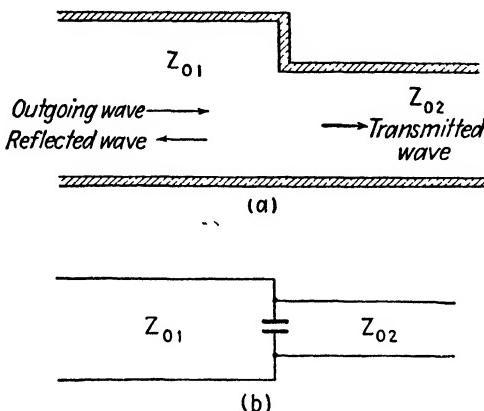
Fig. 4.—Wave guide with a step discontinuity and equivalent circuit.

modes may then appear in the vicinity of the aperture, but they will not exist at any appreciable distance away from the aperture. Under these conditions, the higher-order modes represent reactive energy storage, which is similar to the energy storage in a lumped inductance or capacitance. Thus, a step discontinuity, such as that shown in Fig. 4, may be represented by two transmission lines, having different characteristic impedances, which are joined together with a lumped capacitance at the junction.

If the dimensions of the guide of Fig. 4 are such as to pass higher-order modes, then the reflected and transmitted modes will probably contain the allowed higher-order modes as well as the dominant mode. The step discontinuity then serves to convert energy from the dominant mode (in the outgoing wave) into higher-mode energy (in the reflected and transmitted waves). The equivalent circuit shown in Fig. 4b is not valid if the dimensions of the guide are such as to pass the higher-order modes.

Apertures, such as those shown in Fig. 5, may have either inductive or capacitive characteristics. If the electric-intensity vector is parallel to the aperture sides, as shown in Fig. 5a ($TE_{0,1}$ mode assumed), the currents can

flow vertically in the aperture wall to terminate the magnetic field. The reactive energy is then largely in the magnetic field and the equivalent circuit consists of a transmission line shunted by an inductance.

On the other hand, if the electric-intensity vector is perpendicular to the sides of the aperture, as shown in Fig. 5b, the current flow in the aperture wall is interrupted and an electric field appears across the gap. For this condition, the reactive energy storage is in the electric field and the aperture appears capacitive. The equivalent circuit then consists of a uniform line shunted by a lumped capacitance.
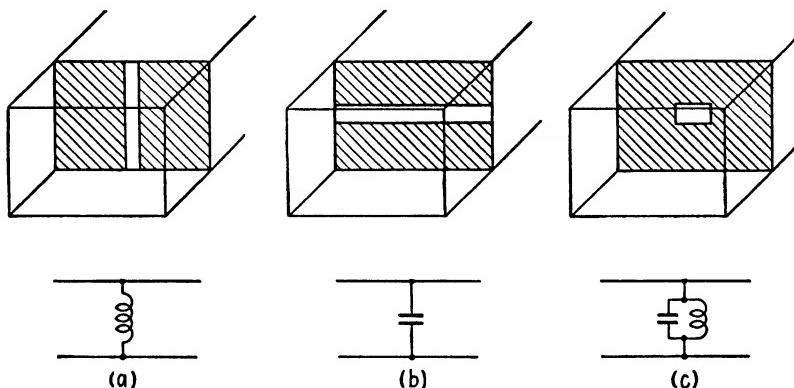
FIG. 5.—Apertures in wave guides and their equivalent circuits (for the dominant mode).

The aperture shown in Fig. 5c may be either capacitive or inductive, depending upon the relative dimensions of the width and height of the aperture. By properly proportioning the aperture, the equivalent inductive and capacitive reactances may be made equal and the equivalent circuit is analogous to an antiresonant parallel *L-C* circuit. Since the impedance of such a circuit is very high, it has very little shunting effect and the wave will therefore pass through the aperture almost as though the aperture were not present.

17.06. Practical Aspects of Resonators.[1]—If a wave guide, excited from a microwave source, is terminated in a conducting end wall, there will be nodal planes of electric intensity at distances from the end wall corresponding to $z = n\lambda_p/2$, where n is any integer and λ_p is the wavelength for the given mode in a direction parallel to the guide walls. A second conducting end wall may be added so as to coincide with the nodal plane of electric intensity without altering the field distribution between the two end walls. It is assumed, of course, that the exciting antenna is in the guide between the two end walls. The totally enclosed guide then becomes a resonator

[1] WILSON, I. G., C. W. SCHRAMM, and J. P. KINZER, High-Q Resonant Cavities for Microwave Testing, *Bell System Tech. J.*, vol. 25, pp. 408–434; July, 1946.

with discrete resonant frequencies and resonance properties somewhat similar to those of the resonant line or the parallel *L-C* circuit.   It is not necessary, however, that a resonator have a simple geometrical shape, such as that described above.   In fact, any closed conducting surface may be considered to be a resonator, regardless of shape.

A given resonator has theoretically an infinite number of resonant modes. Each mode corresponds to a definite resonant frequency (or wavelength). If the exciting source has a frequency differing appreciably from any of the
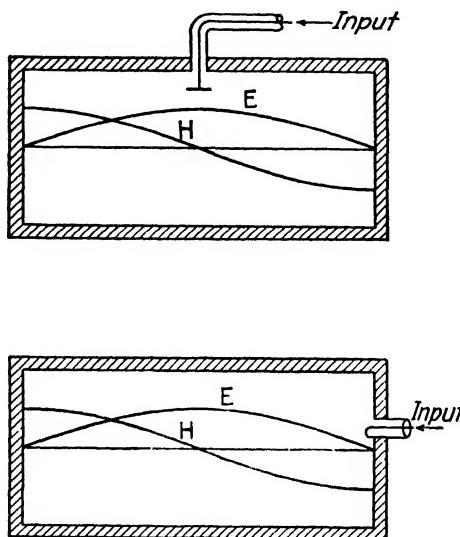
Fig. 6.—Methods of exciting a resonator.

resonant frequencies, the electromagnetic field in the resonator will be extremely small.   However, as the frequency of the impressed signal approaches one of the resonant frequencies, pronounced electromagnetic oscillations appear, as evidenced by relatively large standing waves of electric and magnetic intensity in the resonator.   The maximum amplitude of the standing wave occurs when the frequency of the impressed signal is equal to a resonant frequency.   In general, the various resonant frequencies are not harmonically related, although in exceptional cases they may be harmonic.   Harmonic resonant frequencies are most likely to occur in the rectangular and spherical resonators.   The mode having the lowest resonant frequency (or the longest wavelength) is known as the *dominant mode*.

17.07. Methods of Determining the Resonant Frequencies.—There are a number of ways of evaluating the resonant frequencies of a resonator. One method, which is particularly applicable to resonators of simple geometry, consists of writing the field-intensity equations as outgoing and re-

flected waves in the manner described in Secs. 17.02 and 17.03. The reso-
nant frequencies are then determined by the requirement that the tangential
electric intensities must be zero at the boundary walls (assuming perfectly
conducting walls). Another way of stating this requirement is that the
wave impedance must be zero at the resonator walls.

A useful resonance criterion states that if a pure reactance circuit is
broken into at any junction, such as at *ab* in Fig. 7a or 7b, resonance occurs
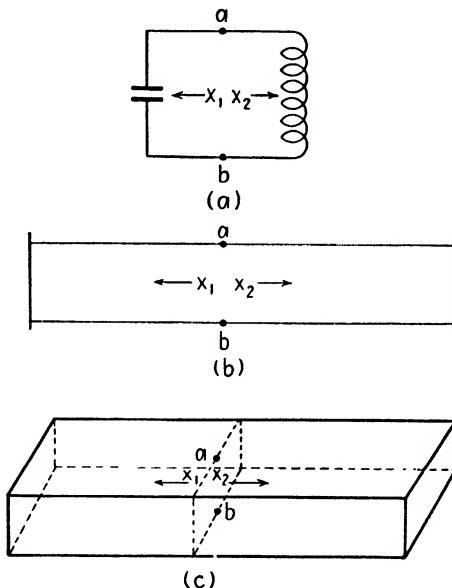
FIG. 7.—Resonance occurs when $X_1 = -X_2$.

at that frequency for which the reactances are equal in magnitude but
opposite in sign, *i.e.*, when $X_1 = -X_2$. This criterion may be applied to
a lossless resonator by writing an expression for the reactance looking both
ways at the plane *ab* in Fig. 7c. Resonance occurs when these reactances
are equal and opposite.

A more general method of determining the resonant frequencies consists
of solving Maxwell's field equations to obtain the field intensities which
satisfy the given boundary conditions. The boundary conditions deter-
mine the resonant frequencies.

Another method is based upon the fact that at resonance the peak energy
storage in the electric and magnetic fields are equal. Consequently, if
expressions can be obtained for the energy storage in the electric and mag-
netic fields, these can be equated to obtain an expression which can be
solved for the resonant frequency.

Finally, certain types of resonators have a construction similar to that of the coaxial line. These resonators may be treated by the transmission-line methods described in Sec. 10.03. In some cases it is possible to use approximate formulas to evaluate the equivalent inductance and capacitance of the resonator. The resonant frequency is then determined by the familiar relationship $\omega_r = 1/\sqrt{LC}$.

A reentrant resonator is one in which the metallic boundaries extend into the interior of the resonator. Several different types of reentrant resonators are shown in Fig. 8. The resonators of Fig. 8a and 8c are
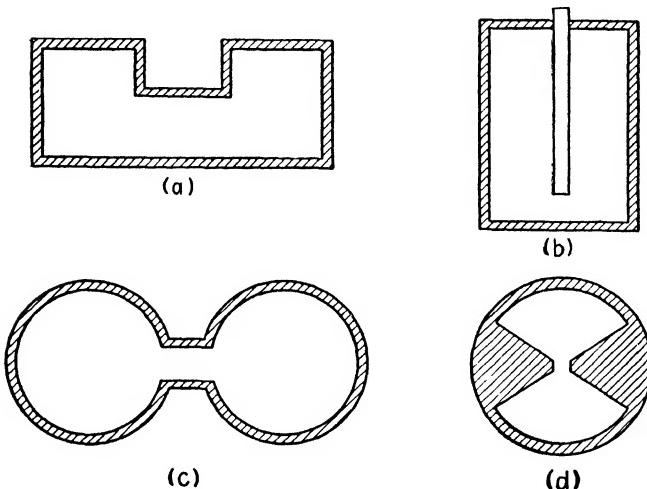
FIG. 8.—Reentrant resonators.

similar to those used in klystron oscillators. Figure 8b shows a resonator which is essentially a coaxial line terminated by a lumped capacitance (the capacitance across the gap between the center conductor and the end wall). The resonator of Fig. 8d is spherical in shape and has conical dimples.

It is sometimes convenient to represent a resonator by an equivalent parallel R-L-C circuit. To obtain the equivalent inductance, the energy storage in the magnetic field is first evaluated. This energy storage is equated to $\frac{1}{2}LI^2$, where L is the equivalent inductance and I is the current in one of the resonator walls. The equivalent capacitance is obtained by evaluating the energy storage in the electric field and equating this to $\frac{1}{2}CV^2$, where V is the voltage difference between two opposite points on the resonator walls, taken where the voltage difference is a maximum.

An equivalent shunt resistance R may be evaluated by expressing the power loss in the resonator as $P_L = V^2/2R$, yielding $R = V^2/2P_L$. If the power loss and voltage are known, the shunt resistance can be determined. The values of R, L, and C in the equivalent circuit of a resonator

are sometimes useful in appraising the over-all merits of the resonator. For example, the same resonant frequency can be obtained with a wide variety of resonator shapes provided that the L-C product is held constant. However, the Q of the resonator increases with L; hence a high L/C ratio results in a high Q.

An interesting and useful principle of similitude states that if two resonators have identical shapes but different scale dimensions, the resonant frequencies of the resonators are inversely proportional to their linear dimensions. If it is desired to construct a resonator having a given shape and resonant frequency, an experimental resonator may first be constructed having the desired shape and approximate size. The resonant frequency may then be measured, and the ratio of the experimental resonant frequency to the desired resonant frequency gives the scale factor to be used in constructing the desired resonator.

17.08. Reactance Method of Determining the Resonant Frequencies.— Consider a rectangular resonator as shown in Fig. 7c, with perfectly conducting walls and a lossless dielectric. At the resonant frequency the reactances X_1 and X_2 looking in opposite directions from a point in the plane ab are equal in magnitude and opposite in sign. Let us now transfer our point of observation to the right-hand wall. The reactance X_2 looking to the right (into the perfectly conducting end wall) is zero. Consequently, at resonance, the reactance X_1 looking to the left must likewise be zero. Applying Eq. (17.03-3), we obtain

$$X_1 = jZ_0 \tan \beta_1 c = 0 \qquad (1)$$

where c is the length of the resonator. Equation (1) is satisfied when $\beta_1 c = \pi p$ or $\beta_1 = p\pi/c$, where p is any positive integer.

Our analysis of wave guides gave the z-directed propagation constant in a lossless rectangular guide as

$$\Gamma = j\beta_1 = \sqrt{k_x^2 + k_y^2 - \omega^2 \mu \epsilon} \qquad (2)$$

This is likewise the propagation constant in the Z direction for the resonator. At the resonant frequency we have from Eq. (1), $\beta_1 = p\pi/c$. Since β_1 is given by an expression similar to the expressions for k_x and k_y, we let $\beta_1 = k_z$, obtaining

$$k_x = \frac{m\pi}{a} \qquad k_y = \frac{n\pi}{b} \qquad k_z = \frac{p\pi}{c} \qquad (3)$$

Inserting k_z for β_1 into Eq. (2) and solving for the resonant frequency f_r, we obtain

$$f_r = \frac{v_c}{2\pi} \sqrt{k_x^2 + k_y^2 + k_z^2} \qquad (4)$$

where $v_c = 1/\sqrt{\mu\epsilon}$. The substitution $\omega_r^2\mu\epsilon = (2\pi/\lambda_r)^2$ yields the resonant wavelength

$$\lambda_r = \frac{2\pi}{\sqrt{k_x^2 + k_y^2 + k_z^2}}$$

$$= \frac{2}{\sqrt{(m/a)^2 + (n/b)^2 + (p/c)^2}} \tag{5}$$

In these equations v_c and λ_r are the velocity and wavelength for the given signal in unbounded dielectric.

Modes in rectangular resonators are designated as either $TE_{m,n,p}$ or $TM_{m,n,p}$ modes. In conformity with the wave-guide terminology, the TE modes are characterized by $E_z = 0$, whereas the TM modes have $H_z = 0$. It is possible, of course, for TE and TM modes to exist simultaneously in the resonator. The integers m, n, and p represent the half-wave periodicity in the x, y, and z directions, respectively. Since Eqs. (4) and (5) are valid for either TE or TM modes, it follows that there are two possible modes for every resonant frequency, one of these being a TE mode and the other a TM mode. An exception occurs when any one of the integers is zero. Modes of the type $TE_{0,n,p}$ can exist, but $TM_{0,n,p}$ modes cannot exist. At least two of the integers m, n, and p must have values greater than zero in order for the field to exist. A single resonant frequency which has two or more modes of oscillation is known as a *degenerate frequency*. In rectangular resonators, all resonant frequencies for which m, n, and p have non-zero values are twofold degenerate.

The lowest resonant frequency, *i.e.*, the frequency of the dominant mode for a rectangular resonator, occurs when the integer associated with the smallest dimension is zero and the other two are unity. Thus if a is the smallest dimension, the $TE_{0,1,1}$ mode is the dominant mode, with

$$f_r = \frac{v_c\sqrt{b^2 + c^2}}{2bc} \tag{6}$$

$$\lambda_r = \frac{2bc}{\sqrt{b^2 + c^2}} \tag{7}$$

For a cube, we have $a = b = c$ and $\lambda_r = \sqrt{2}a$, or the resonant wavelength is equal to the length of the diagonal of one face of the cube.

17.09. Rectangular Resonator—Solution by Maxwell's Equations.— Let us now consider the analysis of the rectangular resonator from the viewpoint of Maxwell's equations. Again we assume that the resonator walls are perfectly conducting and that the dielectric is lossless. We start

the analysis by assuming electric-intensity components which satisfy both the boundary conditions and the divergence equation. These are

$$E_x = E_1 \cos k_x x \sin k_y y \sin k_z z$$

$$E_y = E_2 \sin k_x x \cos k_y y \sin k_z z \qquad (1)$$

$$E_z = E_3 \sin k_x x \sin k_y y \cos k_z z$$

where the time function $e^{j\omega t}$ has been omitted for brevity. The insertion of these into the divergence equation $\nabla \cdot \bar{E} = 0$ yields

$$k_x E_1 + k_y E_2 + k_z E_3 = 0 \qquad (2)$$

This imposes a restriction upon the values of E_1, E_2, and E_3. If we had assumed all sine terms or all cosine terms in Eq. (1), these would not have satisfied the divergence equation and hence such a field could not exist.
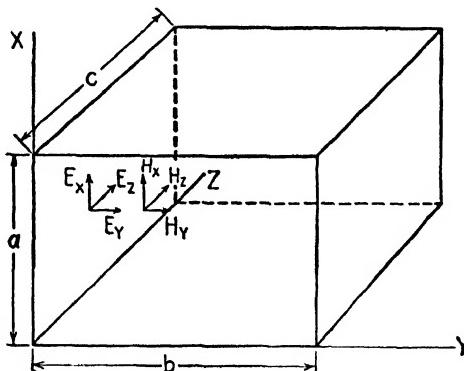
FIG. 9.—The rectangular resonator.

In rectangular coordinates the wave equation may be written for any component, such as $\nabla^2 E_x = \gamma^2 E_x$. Inserting E_x from Eq. (1) into the wave equation, together with $\gamma^2 = -\omega_r^2 \mu \epsilon = -(2\pi/\lambda_r)^2$, we obtain the resonant frequency and wavelength

$$f_r = \frac{v_c}{2\pi} \sqrt{k_x^2 + k_y^2 + k_z^2}. \qquad (3)$$

$$\lambda_r = \frac{2}{\sqrt{(m/a)^2 + (n/b)^2 + (p/c)^2}} \qquad (4$$

These are identical to Eqs. (17.08-4 and 5).

The magnetic intensities may be readily obtained by inserting the electric intensities given by Eq. (1) into the curl equation (13.06-1). This process yields

$$H_x = \frac{(k_z E_2 - k_y E_3)}{j\omega\mu} \sin k_x x \cos k_y y \cos k_z z$$

$$H_y = \frac{(k_x E_3 - k_z E_1)}{j\omega\mu} \cos k_x x \sin k_y y \cos k_z z \qquad (5)$$

$$H_z = \frac{(k_y E_1 - k_x E_2)}{j\omega\mu} \cos k_x x \cos k_y y \sin k_z z$$

For transverse-electric modes, one of the coefficients E_1, E_2, or E_3 must be zero, whereas for transverse magnetic modes, one of the coefficients of the magnetic intensities must be zero.

If any one of the integers m, n, or p is zero, the corresponding k is also zero and two of the components of electric intensity in Eq. (1) vanish, leaving the intensity component which is polarized in the direction of the zero integer axis. For example, the $TE_{0,n,p}$ mode has $E_y = E_z = 0$, and only E_x exists. If two of the integers are zero, all of the field intensity components vanish and consequently there can be no such mode of oscillation.

If we apply the resonance criterion shown in Fig. 7 to a lossless transmission line which is short-circuited at both ends, we find that resonance occurs when $\tan \beta l = 0$, where l is the length of the line. This requires that $\beta l = n\pi$, or $\beta = n\pi/l$, where n is any integer. The resonant wavelength is then

$$\lambda_r = \frac{2l}{n} \qquad (6)$$

Comparison of Eqs. (4) and (6) shows that transmission-line resonance may be regarded as a special case of Eq. (4) in which two of the integers are zero. Hence, resonance on a transmission line may be viewed as an oscillation in one degree of freedom (the longitudinal direction). Resonator oscillations, on the other hand, may have either two or three degrees of freedom. Oscillations in two degrees of freedom occur when one of the integers is zero, and in three degrees of freedom when none of the integers is zero.

The resonant frequencies or wavelengths of a rectangular resonator may be represented by the lattice structure shown in Fig. 10.[1,2] This is obtained

---

[1] This is similar to a method used for determining the resonant frequencies of an acoustical resonator as described by P. M. MORSE in "Vibration and Sound," chap. 8, McGraw-Hill Book Company, Inc., New York, 1936.

[2] CONDON, E. U., Principles of Microwave Radio, *Rev. Modern Phys.*, vol. 15, pp. 341-389; October, 1942.

by writing Eqs. (3) and (4) in the form

$$\left(\frac{2f_r}{v_c}\right)^2 = \left(\frac{2}{\lambda_r}\right)^2 = \left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2 + \left(\frac{p}{c}\right)^2 \tag{7}$$

and plotting the values of m/a, n/b, and p/c along the x, y, and z axes, respectively. Each rectangular box in the lattice corresponds to one set of values of m/a, n/b, and p/c, and the length of its diagonal is proportional to $2f_r/v_c$ or $2/\lambda_r$. If one of the integers is zero, this is represented by a rectangle in one of the coordinate planes and the length of the diagonal of the rectangle is proportional to $2f_r/v_c$ or $2/\lambda_r$.
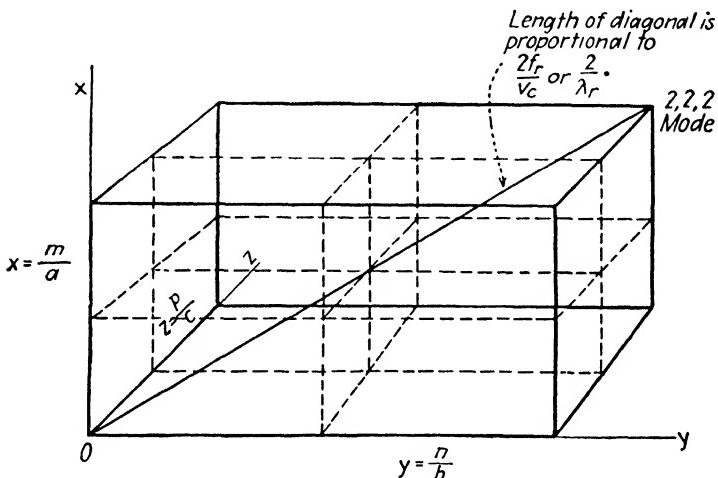
FIG. 10.—Lattice structure representing the resonant frequencies of a rectangular resonator.

The number of resonant frequencies mounts rapidly with increasing values of m, n, and p. For example, there are four resonant frequencies including and below the $TE_{1,1,1}$ mode, and 20 resonant frequencies below its second harmonic, the $TE_{2,2,2}$ mode. Most of the resonant frequencies are nonharmonic. Each rectangular box corresponds to two modes, a TE mode and a TM mode, both having the same resonant frequency.

If two or more resonator dimensions are equal, the integers associated with these dimensions are interchangeable without altering the resonant frequency, hence their resonant frequencies are degenerate.

17.10. *Q* of Resonators.—The Q of a resonant system is a measure of its frequency selectivity. The Q of a resonator may be evaluated by using the definition given by Eq. (10.02-3), *i.e.*,

$$Q = \omega \frac{W_s}{P_L} \tag{10.02-3}$$

where W_s is the peak value of energy storage in the field of the resonator and P_L is the time-average power loss.

In a high-Q resonator, the electric and magnetic intensities are in time quadrature. When the electric intensity has its maximum value, the magnetic intensity is zero, and vice versa. Therefore the peak energy storage in either the electric field or the magnetic field may be used in Eq. (10.02-3). The peak value of energy density stored in the electric and magnetic fields may be expressed as $w_e = \frac{1}{2}\epsilon | E |^2$ and $w_m = \frac{1}{2}\mu | H |^2$, respectively, where $| E |$ and $| H |$ are the peak values of the intensities. The total energy storage in the resonator is obtained by integrating this energy density over the volume of the resonator, or

$$W_s = \frac{\mu}{2} \int_\tau | H |^2 \, d\tau = \frac{\epsilon}{2} \int_\tau | E |^2 \, d\tau \tag{1}$$

The time-average power loss is evaluated by integrating the power density given by Eq. (14.04-4) over the inside surface of the resonator, thus

$$P_L = \frac{\mathscr{R}_s}{2} \int_s | H_t |^2 \, ds \tag{2}$$

where $\mathscr{R}_s$ is the skin-effect resistance per unit area of conductor, as given by Eq. (14.07-4), and H_t is the peak value of the tangential magnetic intensity.

Inserting these expressions for W_s and P_L into Eq. (10.02-3), we obtain an expression for the Q of the resonator

$$Q = \frac{\omega\mu \int_\tau | H |^2 \, d\tau}{\mathscr{R}_s \int_s | H_t |^2 \, ds} \tag{3}$$

The numerator of Eq. (3) is integrated over the volume of the resonator, while the denominator is integrated over the inside surface area of the resonator walls.

An approximate expression for the Q of a rectangular resonator may be obtained by assuming that the standing wave of electric intensity has a sinusoidal distribution. The value of $| H_t |^2$ at the resonator walls is approximately twice the value of $| H |^2$ averaged over the volume. Hence, Eq. (3) may be represented approximately by

$$Q = \frac{\omega\mu\tau}{2\mathscr{R}_s s} \tag{4}$$

where r is the volume of the resonator and s is the inside surface area of the resonator walls. Hence the Q is roughly proportional to the ratio of volume to surface area of the resonator.

Extremely high Q's are attainable with well-designed resonators. Typical values of Q for unloaded resonators are from 2,000 to 100,000. Silver and gold plating are often used to reduce the skin-effect resistance and thereby increase the Q. If sliding pistons are used to tune the resonator, the loss due to the current flow across the contact resistance between the piston and the resonator walls results in an appreciable lowering of the Q of the resonator. The use of spring contact fingers of low-resistance material on the piston helps to reduce this loss.

If there are losses in the dielectric of the resonator as well as in the resonator walls, the Q for each type of loss may be computed separately. Representing these Q's by Q_1 and Q_2, the effective Q of the resonator is

$$Q = \frac{Q_1 Q_2}{Q_1 + Q_2} \tag{5}$$

The resonant frequency of a resonator varies inversely as the square root of the dielectric constant. It is therefore possible to reduce the size of a resonator for a given resonant frequency by the use of a dielectric having a dielectric constant greater than unity. However, all known dielectric materials have appreciable losses at the microwave frequencies, hence their use in resonators results in a substantial decrease in the effective Q of the resonator.

17.11. The Q of a Rectangular Resonator.—The Q of a resonator may be evaluated by the methods of Sec. 17.10. To illustrate the method, the Q will be determined for a rectangular resonator operating in the $TE_{0,n,p}$ mode. The intensities are obtained by setting $k_x = 0$ in Eqs. (17.09-1 and 5), giving

$$E_x = E_1 \sin k_y y \sin k_z z$$

$$H_y = -\frac{k_z E_1}{j\omega_r \mu} \sin k_y y \cos k_z z \tag{1}$$

$$H_z = \frac{k_y E_1}{j\omega_r \mu} \cos k_y y \sin k_z z$$

The peak energy storage in the electric field is obtained by inserting E_x into Eq. (17.10-1) and integrating, yielding

$$W_s = \frac{a\epsilon E_1^2}{2} \int_0^c \int_0^b \sin^2 k_y y \sin^2 k_z z \, dy \, dz$$

$$= \epsilon E_1^2 \frac{abc}{8} \tag{2}$$

The time-average power loss is obtained by inserting the tangential components of magnetic intensity into Eq. (17.10-2) and integrating. Referring to Fig. 9 and Eq. (1), the tangential magnetic intensities are

at side walls

$$\left| H_{zt} \right| = \frac{k_y E_1}{\omega_r \mu} \sin k_z z$$

at front and back walls

$$\left| H_{yt} \right| = \frac{k_z E_1}{\omega_r \mu} \sin k_y y$$

at top and bottom walls

$$\left| H_{yt} \right| = \frac{k_z E_1}{\omega_r \mu} \sin k_y y \cos k_z z$$

$$\left| H_{zt} \right| = \frac{k_y E_1}{\omega_r \mu} \cos k_y y \sin k_z z$$

The power loss then becomes

$$P_L = \mathcal{R}_s \left(\frac{E_1}{\omega_r \mu} \right)^2 \left[k_y^2 \int_0^c \int_0^a \sin^2 k_z z \, dx \, dz + k_z^2 \int_0^b \int_0^a \sin^2 k_y y \, dx \, dy \right.$$

$$\left. + k_z^2 \int_0^c \int_0^b \sin^2 k_y y \cos^2 k_z z \, dy \, dz + k_y^2 \int_0^c \int_0^b \cos^2 k_y y \sin^2 k_z z \, dy \, dz \right]$$

Performing the indicated operations and substituting the values of k_x, k_y, and k_z from Eq. (17.08-3), we obtain

$$P_L = \frac{\mathcal{R}_s}{2} \left(\frac{\pi E_1}{\omega_r \mu} \right)^2 \left[\frac{n^2 ac}{b^2} + \frac{p^2 ab}{c^2} + \frac{bc}{2} \left(\frac{p^2}{c^2} + \frac{n^2}{b^2} \right) \right] \tag{3}$$

Inserting W_s and P_L into Eq. (10.02-3), with the substitutions $\omega_r^2 \mu \epsilon = (2\pi/\lambda_r)^2$ and $\eta = \sqrt{\mu/\epsilon}$, gives

$$Q = \frac{2\pi \eta abc}{\mathcal{R}_s \lambda_r^3 \{ (n^2 ac/b^2) + (p^2 ab/c^2) + bc/2[(p^2/c^2) + (n^2/b^2)] \}} \tag{4}$$

For a resonator in which $b = c$, this becomes

$$Q = \frac{2\pi \eta ab^2}{\mathcal{R}_s \lambda_r^3 \{ [(a/b) + \frac{1}{2}](n^2 + p^2) \}} \tag{5}$$

and for the cube $a = b = c$,

$$Q = \frac{4\pi \eta a^3}{3 \mathcal{R}_s \lambda_r^3 (n^2 + p^2)} \tag{6}$$

If the cubical resonator is excited in the $TE_{0,1,1}$ mode, we have $\lambda_r = \sqrt{2} a$ and $Q = 0.741 \eta / \mathcal{R}_s$.

17.12. Cylindrical Resonator.—The cylindrical resonator having a circular cross section may be analyzed by the methods of Sec. 17.08. The intensities of the outgoing and reflected waves and the wave impedance of a circular guide with a perfectly conducting end wall may be expressed by equations similar to Eqs. (17.03-1 to 3). Thus, for E_ρ and H_ϕ we may write

$$E_\rho = jH_{\phi R}\, Z_{01} \sin \beta_1 z\, f(\rho, \phi) \tag{1}$$

$$H_\phi = H_{\phi R} \cos \beta_1 z\, f(\rho, \phi) \tag{2}$$

$$Z = \frac{E_\rho}{H_\phi} = jZ_{01} \tan \beta_1 z \tag{3}$$

where $H_{\phi R}$ is the value of H_ϕ at the end wall, and the function $f(\rho, \phi)$ may be obtained from Eqs. (16.06-15 or 16). Similar expressions may be written for the E_ϕ, H_ρ pair, the impedance being the same as that given by Eq. (3).

Resonance occurs when the second conducting end wall is at a point of zero impedance. This requires that $\tan \beta_1 c = 0$, where c is the axial length of the resonator. This requirement is satisfied by $\beta_1 c = p\pi$ or $\beta_1 = p\pi/c$, where p is any integer. An expression for β_1 in circular guides or resonators is obtained by setting $\Gamma = j\beta_1$ in Eq. (16.06-19), yielding

$$\beta_1 = \sqrt{\omega_r^2 \mu\epsilon - \frac{(kb)^2}{b^2}} \tag{4}$$

The substitutions $\beta_1 = p\pi/c$ and $\omega_r^2\mu\epsilon = (2\pi/\lambda_r)^2$ yield the resonant frequency and wavelength,

$$f_r = \frac{v_c}{2\pi} \sqrt{\left(\frac{p\pi}{c}\right)^2 + \frac{(kb)^2}{b^2}} \tag{5}$$

$$\lambda_r = \frac{2\pi}{\sqrt{(p\pi/c)^2 + (kb)^2/b^2}} \tag{6}$$

The values of (kb) are the roots of $J_n(kb) = 0$ for *TM* modes and of $J'_n(kb) = 0$ for *TE* modes, as given in Table 5, Chap. 16. In circular resonators the modes are designated $TE_{n,m,p}$ and $TM_{n,m,p}$. The integer n determines the periodicity in the ϕ direction [see Eqs. (16.06-15 and 16)], m denotes the number of zeros of electric intensity in the radial direction (exclusive of the zero on the axis), and p is the number of half wavelengths in the axial direction. Since the values of (kb) are different for *TE* and *TM* modes, the resonant frequencies do not have the type of degeneracy found in rectangular resonators.

For $n = 0$ the electric intensity has a radial distribution corresponding to a zero-order Bessel function and there is no variation of the field in the ϕ direction. For *TM* modes we may also have $p = 0$, corresponding to

uniform intensity in the axial direction. The resonant wavelength of $TM_{0,m,0}$ modes is

$$\lambda_r = \frac{2\pi b}{(kb)} \tag{7}$$

The $TM_{0,1,0}$ mode has a value $(kb) = 2.405$ and a resonant wavelength of $\lambda_r = 2.61b$. For the $TM_{0,2,0}$ mode we have $(kb) = 5.520$ and $\lambda_r = 1.14b$.

For TE modes n may be zero but p must have nonzero integer values. The $TE_{1,1,1}$ mode has a value of $(kb) = 1.84$ and

$$\lambda_r = \frac{2\pi}{\sqrt{(\pi/c)^2 + (3.38/b^2)}}$$

The resonant wavelength of the $TE_{1,1,1}$ mode is less than that of the $TM_{0,1,0}$ mode if $c < 2.02b$.

The $TE_{0,m,p}$ modes in circular resonators are unique in that no axial currents flow in the side walls of the resonator and there are no currents circulating between the end walls and side walls for these modes. Consequently, resonators operating in any one of these modes may be tuned by means of a sliding piston without appreciably lowering the Q. The piston may be loose fitting in order to discourage undesired modes. The $TE_{0,1,1}$ mode has a value of $(kb) = 3.832$ and $\lambda_r = 2\pi/\sqrt{(\pi/c)^2 + (14.70/b^2)}$.

17.13. Q of the Cylindrical Resonator.—The Q of the cylindrical resonator will be derived for the $TM_{0,m,0}$ mode. This mode has a radial variation of intensity corresponding to the zero-order Bessel function, no circumferential variation, and no variation axially. The intensities, as obtained from Eqs. (16.06-16), are

$$E_\rho = -j\frac{\beta_1}{k} E_0 J_0'(k\rho) \tag{1}$$

$$E_z = E_0 J_0(k\rho) \tag{2}$$

$$H_\phi = -j\frac{\omega\epsilon}{k} E_0 J_0'(k\rho) \tag{3}$$

From Eq. (15.10-4), we obtain $J_0'(k\rho) = -J_1(k\rho)$, and for convenience we write H_ϕ as $H_\phi = H_0 J_1(k\rho)$. The peak energy storage in the magnetic field is obtained by inserting H_ϕ into Eq. (17.10-1), yielding

$$W_s = \frac{\mu H_0^2 c}{2} \int_0^b 2\pi\rho J_1^2(k\rho) \, d\rho \tag{4}$$

The integration is given by Eq. (15.10-14). Substituting

$$J_0'(kb) = -J_1(kb)$$

and remembering that $J_0(kb) = 0$ for TM modes, we obtain

$$\int_0^b \rho J_1^2(k\rho) \, d\rho = \frac{b^2}{2} J_1^2(kb)$$

Equation (4) then becomes

$$W_s = \frac{\pi\mu H_0^2 c b^2}{2} J_1^2(kb) \tag{5}$$

The power loss is obtained by inserting $H_{\phi t}$ from Eq. (3) into (17.10-2), giving

$$P_L = \frac{\mathcal{R}_s H_0^2}{2} \left[2\pi b c J_1^2(kb) + 2\int_0^b 2\pi\rho J_1^2(k\rho)\, d\rho\right]$$

$$= \pi b \mathcal{R}_s H_0^2 J_1^2(kb)(c + b) \tag{6}$$

The Q for the $TM_{0,m,0}$ modes then becomes

$$Q = \frac{\omega_r W_s}{P_L} = \frac{\pi\eta bc}{\lambda_r \mathcal{R}_s(c + b)} \tag{7}$$

A similar derivation for the $TM_{0,m,p}$ modes yields

$$Q = \frac{\pi\eta bc}{\lambda_r \mathcal{R}_s(c + 2b)} \tag{8}$$

17.14. The Spherical Resonator.—Since the sphere has the highest ratio of volume to surface area, it offers attractive possibilities as a high-Q resonator. In the following analysis we shall consider the natural modes of oscillation inside of a perfectly conducting spherical shell with a lossless dielectric. The analysis of the oscillating sphere has been treated by Debye,[1] Stratton,[2] Condon,[3] and others. The complete analysis is beyond the scope of this text and therefore we shall draw upon the treatment given by Condon.

The vector intensities in the resonator may be constructed from a scalar wave function U which is similar to the solution of the scalar wave equation given by Eq. (15.12-21), except that the expression for U is multiplied by (kr).

$$U = [(kr)j_n(kr)]P_n^m(\cos\theta)[A\cos m\phi + B\sin m\phi] \tag{1}$$

where $j_n(kr)$ is the spherical Bessel function defined by Eq. (15.08-1) and $P_n^m(\cos\theta)$ is the associated Legendre function described in Sec. 15.12.

For brevity, we let $Y_{nm} = P_n^m(\cos\theta)(A\cos m\phi + B\sin m\phi)$ and Eq. (1) then becomes

$$U = (kr)j_n(kr)Y_{nm} \tag{2}$$

[1] DEBYE, Der Lichtdruck auf kugeln von beliebigem Material, *Ann. Physik*, vol. 30, p. 57; 1909.

[2] STRATTON, J. A., "Electromagnetic Theory," chaps. 7 and 9, McGraw-Hill Book Company, Inc., New York, 1941.

[3] CONDON, E. U., Principles of Microwave Radio, *Rev. Modern Phys.*, vol. 14, pp. 341–389; October, 1942.

In spherical systems the TE mode is characterized by $E_r = 0$ and the TM modes by $H_r = 0$. Condon has shown that the following equations express the intensities in terms of the scalar wave function,

<center><u>TE modes</u></center>

$$E_r = 0$$

$$E_\theta = -\frac{j\omega\mu}{r\sin\theta}\frac{\partial U}{\partial\phi}$$

$$E_\phi = \frac{j\omega\mu}{r}\frac{\partial U}{\partial\theta}$$

$$H_r = k^2 U + \frac{\partial^2 U}{\partial r^2}$$

$$H_\theta = \frac{1}{r}\frac{\partial^2 U}{\partial r\partial\theta}$$

$$H_\phi = \frac{1}{r\sin\theta}\frac{\partial^2 U}{\partial r\,\partial\phi}$$

$$\left.\right\}\quad(3)$$

<center><u>TM modes</u></center>

$$E_r = k^2 U + \frac{\partial^2 U}{\partial r^2}$$

$$E_\theta = \frac{1}{r}\frac{\partial^2 U}{\partial r\,\partial\theta}$$

$$E_\phi = \frac{1}{r\sin\theta}\frac{\partial^2 U}{\partial r\,\partial\phi}$$

$$H_r = 0$$

$$H_\theta = \frac{j\omega\epsilon}{r\sin\theta}\frac{\partial U}{\partial\phi}$$

$$H_\phi = -\frac{j\omega\epsilon}{r}\frac{\partial U}{\partial\theta}$$

$$\left.\right\}\quad(4)$$

In these equations

$$k^2 = \omega^2\mu\epsilon = \left(\frac{2\pi}{\lambda}\right)^2 \tag{5}$$

Inserting U from Eq. (2) into (3) and (4), and remembering that Y_{nm} is a function of θ and ϕ, the intensities become

<center><u>TE modes</u></center>

$$E_r = 0$$

$$E_\theta = -\frac{j\omega\mu}{r\sin\theta}[(kr)j_n(kr)]\frac{\partial Y_{nm}}{\partial\phi}$$

$$E_\phi = \frac{j\omega\mu}{r}[(kr)j_n(kr)]\frac{\partial Y_{nm}}{\partial\theta}$$

$$H_r = \frac{k^2}{r^2}n(n+1)[(kr)j_n(kr)Y_{nm}]$$

$$H_\theta = \frac{k}{r}\frac{\partial}{\partial(kr)}[(kr)j_n(kr)]\frac{\partial Y_{nm}}{\partial\theta}$$

$$H_\phi = \frac{k}{r\sin\theta}\frac{\partial}{\partial(kr)}[(kr)j_n(kr)]\frac{\partial Y_{nm}}{\partial\phi}$$

$$\left.\right\}\quad(6)$$

$$\underline{TM \text{ modes}}$$

$$E_r = \frac{k^2}{r^2} n(n + 1)[(kr)j_n(kr)]Y_{nm}$$

$$E_\theta = \frac{k}{r} \frac{\partial}{\partial(kr)} [(kr)j_n(kr)] \frac{\partial Y_{nm}}{\partial \theta}$$

$$E_\phi = \frac{k}{r \sin \theta} \frac{\partial}{\partial(kr)} [(kr)j_n(kr)] \frac{\partial Y_{nm}}{\partial \phi}$$

$$H_r = 0 \qquad\qquad\qquad\qquad\qquad\qquad (7)$$

$$H_\theta = -\frac{j\omega\epsilon}{r \sin \theta} [(kr)j_n(kr)] \frac{\partial Y_{nm}}{\partial \phi}$$

$$H_\phi = -\frac{j\omega\epsilon}{r} [(kr)j_n(kr)] \frac{\partial Y_{nm}}{\partial \theta}$$

To satisfy the boundary conditions it is necessary that $E_\theta = E_\phi = 0$ at the surface where $r = a$. This requires that

$$j_n(ka) = 0 \qquad TE \text{ modes} \qquad (8)$$

$$\frac{\partial}{\partial(kr)} [(kr)j_n(kr)] \bigg|_{r=a} = 0 \qquad TM \text{ modes} \qquad (9)$$

The roots of Eqs. (8) and (9) determine the values of the resonant frequencies. Equation (5) may be written

$$f_r = \frac{v_c}{2\pi} \frac{(ka)}{a} \qquad (10)$$

$$\lambda_r = \frac{2\pi a}{(ka)} \qquad (11)$$

where the values of (ka) are determined by Eqs. (8) and (9).

We shall designate spherical resonator modes by the symbols $TE_{n,p,m}$ and $TM_{n,p,m}$ where n is the order of the spherical Bessel function, p is an integer denoting the rank of the roots of Eq. (8) or (9) for a given value of n, and m is the periodicity in the ϕ direction. It is interesting to note that the resonant frequency is independent of the integer m. However, in order for the field to exist, it is necessary that $m \leq n$ since $P_n^m(\cos \theta)$ vanishes when m is greater than n. Hence, for a given value of n the integer m may have values from 0 to n inclusive, each corresponding to a separate mode but all having identically the same resonant frequency. Furthermore, the intensity distribution in the ϕ direction may be of the form

cos $m\phi$ or sin $m\phi$ or any linear combination of the two. A spherical resonator, therefore, may oscillate in a number of different modes having the same resonant frequency.

The field vanishes if $n = 0$; hence this does not correspond to an allowed mode. The lowest resonant frequency for either the TE or TM modes

(a)- $TE_{1,1,0}$ Mode   (b)- $TM_{1,1,0}$ Mode
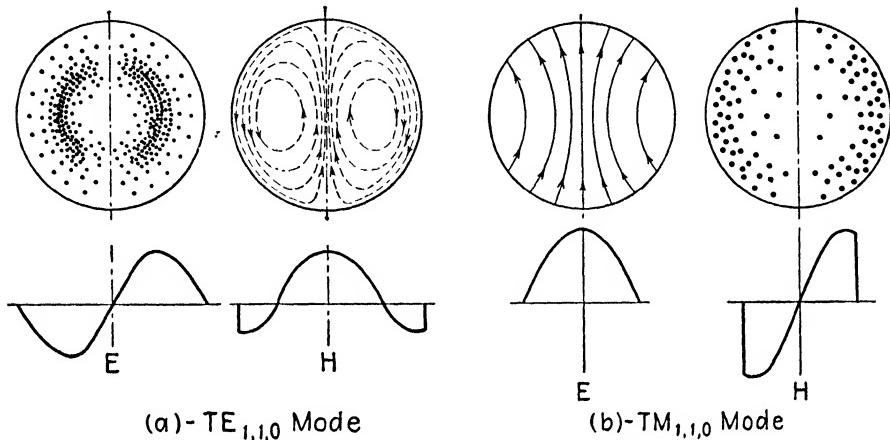
FIG. 11.—Field patterns in the meridian plane of a spherical resonator.

occurs when $n = p = 1$. The corresponding roots of Eqs. (8) and (9) are $(ka)_{1,1} = 4.49$ for the TE mode and $(ka)_{1,1} = 2.75$ for the TM mode. Inserting these into Eq. (11) gives

$$TE_{1,1,m} \qquad \lambda_r = 1.40a \tag{12}$$

$$TM_{1,1,m} \qquad \lambda_r = 2.29a \tag{13}$$

If $m = 0$, the field has no variation in the ϕ direction since the term $A \cos m\phi + B \sin m\phi$ reduces to a constant. The field is then circularly symmetrical and the function $P_n^m(\cos\theta)$ then reduces to the Legendre function $P_n(\cos\theta)$. We then have $Y_{nm} = P_n(\cos\theta)$ and $\partial/\partial\phi = 0$. The intensities given in Eqs. (3) and (4) then simplify to

<u>TE modes</u>

$$\left. \begin{aligned} E_r &= E_\theta = H_\phi = 0 \\ E_\phi &= \frac{j\omega\mu}{r}\left[(kr)j_n(kr)\frac{\partial P_n(\cos\theta)}{\partial\theta}\right] \\ H_r &= \frac{k^2}{r^2}\,n(n+1)[(kr)j_n(kr)]P_n(\cos\theta) \\ H_\theta &= \frac{k}{r}\frac{\partial}{\partial(kr)}[(kr)j_n(kr)]\frac{\partial P_n(\cos\theta)}{\partial\theta} \end{aligned} \right\} \tag{14}$$

TM modes

$$E_\phi = H_r = H_\theta = 0$$

$$E_r = \frac{k^2}{r^2} n(n+1)[(kr)j_n(kr)]P_n(\cos\theta)$$

$$E_\theta = \frac{k}{r} \frac{\partial}{\partial(kr)}[(kr)j_n(kr)]\frac{\partial P_n(\cos\theta)}{\partial\theta} \qquad (15)$$

$$H_\phi = -\frac{j\omega\epsilon}{r}[(kr)j_n(kr)]\frac{\partial P_n(\cos\theta)}{\partial\theta}$$

17.15. *TM*$_{1,1,0}$ Mode in the Spherical Resonator.—As an example, consider the *TM*$_{1,1,0}$ mode. Equation (15.08-2) gives the spherical Bessel function as $j_1(kr) = \frac{1}{kr}\left(\frac{\sin kr}{kr} - \cos kr\right)$. From Eq. (15.12-15) we obtain $P_1(\cos\theta) = \cos\theta$. Inserting these into Eq. (17.14-15), we obtain the intensities

$$E_\phi = H_r = H_\theta = 0$$

$$E_r = \frac{8\pi^2}{\lambda^2 r^2}\left(\frac{\sin kr}{kr} - \cos kr\right)\cos\theta$$

$$E_\theta = -\frac{2\pi}{\lambda r}\left\{\frac{\cos kr}{kr} + \left[1 - \frac{1}{(kr)^2}\right]\sin kr\right\}\sin\theta \qquad (1)$$

$$H_\phi = \frac{j2\pi}{\lambda\eta r}\left(\frac{\sin kr}{kr} - \cos kr\right)\sin\theta$$

To evaluate the energy storage in the resonator we write H_ϕ in the form $H_\phi = Aj_1(kr)\sin\theta$. Inserting this into Eq. (17.10-1) the energy storage in the resonator becomes

$$W_s = \frac{\mu A^2}{2k^3}\int_0^\pi\int_0^a j_1^2(kr)2\pi(kr)^2\sin^3\theta\, d(kr)\, d\theta$$

$$= \frac{4\pi\mu A^2}{3k^3}\left\{\frac{(ka)^3}{2}[j_1^2(ka) - j_0(ka)j_2(ka)]\right\} \qquad (2)$$

The time-average power loss is evaluated by inserting $H_{\phi t} = Aj_1(ka)\sin\theta$ into Eq. (17.10-2), yielding

$$P_L = \frac{\mathcal{R}_s A^2}{2}j_1^2(ka)\int_0^\pi 2\pi a^2\sin^3\theta\, d\theta$$

$$= \frac{4\pi}{3}a^2 A^2\mathcal{R}_s j_1^2(ka) \qquad (3)$$

The resonator Q is obtained by inserting these into $Q = \omega_r W_s / P_L$. With the help of $\omega_r = 2\pi v_c / \lambda_r$ and $a = \lambda_r / 2.29$ from Eq. 17.14-13, we obtain

$$Q = \frac{1.37\eta}{\mathcal{R}_s} \left[1 - \frac{j_0(ka)j_2(ka)}{j_1^2(ka)} \right] \tag{4}$$

The spherical Bessel functions may be evaluated using Eq. (15.08-2). For the $TM_{1,1,0}$ mode we have $(ka)_{1,1} = 2.75$ and $j_0(ka) = 0.139$, $j_1(ka) = 0.386$, and $j_2(ka) = 0.282$. Equation (4) then becomes

$$Q = \frac{1.01\eta}{\mathcal{R}_s} \tag{5}$$

As an example, a silver-plated spherical resonator operating in the $TM_{1,1,0}$ mode with a resonant frequency of 3×10^9 cycles per second would have $\mathcal{R}_s = \sqrt{\omega\mu_2/2\sigma_2} = 0.0139$ and a theoretical Q of 27,400.

Figure 11 shows the electric and magnetic field distributions of the $TE_{1,1,0}$ and $TM_{1,1,0}$ modes in a spherical resonator. The electric and magnetic intensities for either mode are in time quadrature. The current flow in the resonator walls is proportional to $H_{\phi t}$ and flows in a direction perpendicular to $H_{\phi t}$. For the TM mode, the magnetic lines are parallel to the equator and hence the current flows along the meridian lines from pole to pole.

17.16. Orthogonality of Modes.—If two or more modes exist simultaneously in a lossless wave guide or resonator, the resultant field may be expressed as the summation of the fields due to the individual modes. Conversely, any given resultant field may be analyzed by the methods of Fourier analysis to determine the various modes which, when superimposed, give the resultant field. Wave guide and resonator modes have an important property of orthogonality which gives them a degree of independence.

Consider a lossless wave guide having two different modes, represented by p and q. Let the transverse electric intensities of the p and q modes be given by

$$\begin{aligned} E_p &= E_{0p}f(x, y) \\ E_q &= E_{0q}g(x, y) \end{aligned} \tag{1}$$

where the transverse spatial distribution factors are represented by $f(x, y)$ and $g(x, y)$. The condition of orthogonality may be stated mathematically as

$$\int_A f(x, y)g(x, y)\, da = 0 \tag{2}$$

where the integration is over the cross section of the guide. Equation (2) is valid for all modes in lossless guides where $p \neq q$, hence wave-guide modes are orthogonal functions.

The significance of the orthogonality principle becomes apparent when we consider power flow in the guide. The longitudinal power flow for

either mode taken separately, for example the p mode, may be expressed by Eq. (16.10-2), thus $P_T = \dfrac{1}{2Z_0} \displaystyle\int_A |E|^2 \, da = \dfrac{E_{0p}^2}{2Z_0} \displaystyle\int_A [f(x, y)]^2 \, da$, which we assume to have nonzero value. Let us suppose, however, that we were to attempt to compute the power flow by the Poynting-vector method using the transverse electric intensity from the p mode and the transverse magnetic intensity from the q mode. We would then encounter an integral of the type given by Eq. (2), which has zero value. Hence, by virtue of the orthogonality property, the various modes transmit power independently of each other in a lossless guide.

Two resonator modes are orthogonal if

$$\int_\tau f(x, y, z) g(x, y, z) \, d\tau = 0 \tag{3}$$

where the integration is over the volume of the resonator.

PROBLEMS

1. A portion of a rectangular wave guide is hollow and the remaining portion is filled with polystyrene ($\epsilon_r = 2.5$), as shown in Fig. 1. The guide is assumed to be infinitely long and the cross-sectional dimensions are $a = 4$ cm and $b = 6$ cm. The frequency is 3,000 mc per sec. Compute the following for a $TE_{0,1}$ mode:
   (a) The characteristic wave impedances in the hollow and dielectric-filled portions of the guide.
   (b) The reflection and transmission coefficients in the guide.
   (c) The standing-wave ratio.
   (d) The ratio of the power density of the transmitted wave to that of the incident wave.
2. A shielded radio room has dimensions 12 by 12 by 7 ft. Compute the six lowest resonant frequencies of the room as a resonator. Sketch a lattice structure, such as that shown in Fig. 10, to represent the resonant modes of this room. Evaluate the Q at the lowest resonant frequency, assuming that the walls are of copper.
3. Show that the peak energy storage in the electric and magnetic fields of a rectangular resonator are equal at resonance.
4. What are the dimensions of cylindrical resonators to oscillate in the $TM_{1,1,1}$ mode at a frequency of 4,500 mc per sec, assuming
   (a) a hollow resonator?
   (b) a dielectric filled resonator having $\epsilon_r = 5$?
5. Discuss the nature of the fields and the boundary conditions in a resonator which is excited at a frequency other than the resonant frequency.
6. A cylindrical resonator is to be resonant in the $TE_{0,1,1}$ mode at a frequency of 3,000 mc per sec.
   (a) Specify the dimensions of the resonator, assuming that the diameter is equal to the height.
   (b) Write the equations for the field intensities.
   (c) Derive an expression for the Q of the resonator.
   (d) Evaluate the Q, assuming that the walls are silver plated.

CHAPTER 18

APPLICATIONS OF WAVE GUIDES AND RESONATORS

In constructing wave-guide systems, it is desirable to have available a number of devices which are the counterpart of our low-frequency networks. Such devices include impedance transformers, filters, attenuators, bridges, etc. It is also necessary to have accurate measuring devices in order to obtain quantitative data on the electrical performance of systems. In this chapter we shall consider some of the practical aspects of wave guides, as well as a number of useful devices which have found widespread application in wave-guide systems.[1,2,3]

18.01. Methods of Exciting Wave Guides.—The antenna systems for launching various modes in rectangular and circular wave guides are shown in Figs. 1 and 2, respectively. In general, straight-wire antennas are placed so as to coincide with the positions of maximum electric intensity for the desired mode. The loop antenna is placed so as to have a maximum number of magnetic flux linkages for the desired mode. If two or more antennas are used, care must be taken to assure the proper phase relationships between the currents in the various antennas. This may be accomplished by inserting additional lengths of transmission line in one or more of the antenna feeders. Impedance matching may be accomplished by varying the position and depth of the antenna in the guide as well as by the use of impedance matching stubs on the coaxial line feeding the wave guide.

An arrangement for launching a $TE_{0,1}$ mode in one direction only is shown in Fig. 3. This consists of two antennas spaced a quarter wavelength apart and phased in time quadrature. Phasing is accomplished by means of the additional quarter-wavelength section of line in the feed to one of the antennas. The fields radiated by the two antennas are in phase opposition to the left of the antennas and hence cancel each other; whereas in the region to the right of the antennas the fields are in time phase and reinforce each other. The resulting wave therefore travels to the right in the guide.

[1] KEMP, J., Wave Guides in Electrical Communication, *J.I.E.E.* (London), vol. 90, Part III, pp. 90–114; September, 1943.

[2] GAFFNEY, F. J., Microwave Measurements and Test Equipment, *Proc. I.R.E.*, vol. 34, pp. 775–793; October, 1946.

[3] GREEN, E. I., H. J. FISHER, and J. G. FERGUSON, Techniques and Facilities for Microwave Testing, *Bell System Tech. J.*, vol. 25, pp. 435–482; July, 1946.

Two methods of exciting wave guides from coaxial lines are shown in Fig. 4. In Fig. 4a, a relatively large magnetic field exists in the vicinity of the short-circuited termination of the coaxial line. Part of this field extends into the wave guide through the aperture and serves to excite the guide. If the wave-guide dimensions are such as to transmit the dominant
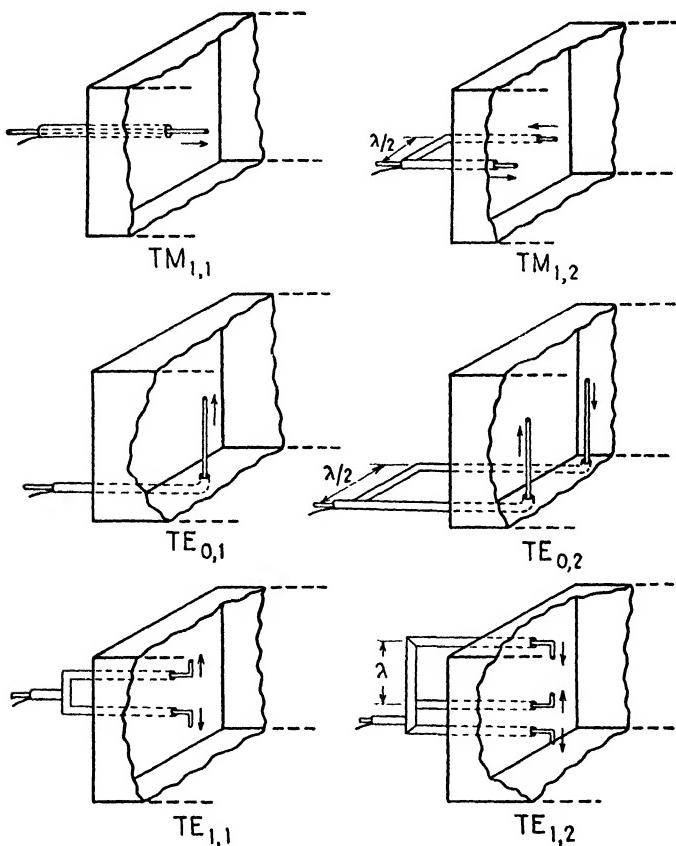
FIG. 1.—Methods of exciting various modes in rectangular wave guides.

mode but attenuate all higher modes, only the dominant mode will exist in the guide at distances exceeding several wavelengths from the aperture. In Fig. 4b, the wave guide begins where the coaxial line leaves off and the field configuration transforms from the principal mode in the coaxial line to the dominant mode in the wave guide (assuming that the higher modes are attenuated).

Figure 5 shows a bridge arrangement of wave guides which may be used to feed two microwave sources into a single guide without introducing coupling between the sources. This device consists of a ring-shaped wave
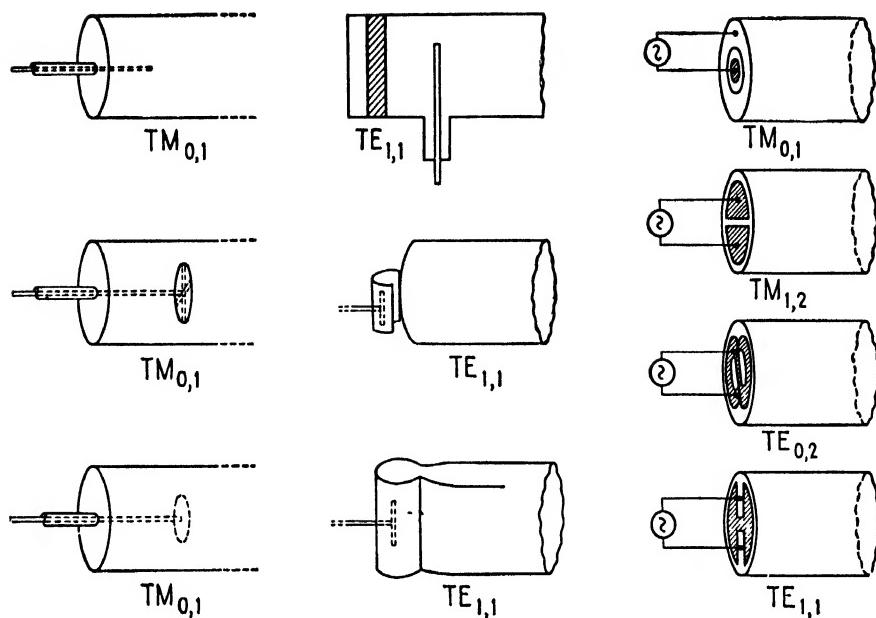
Fig. 2.—Methods of exciting various modes in circular wave guides.
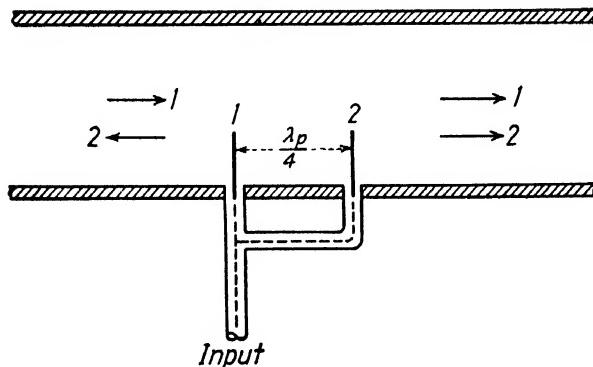
Fig. 3.—A method of launching a $TE_{0,1}$ mode in one direction only.
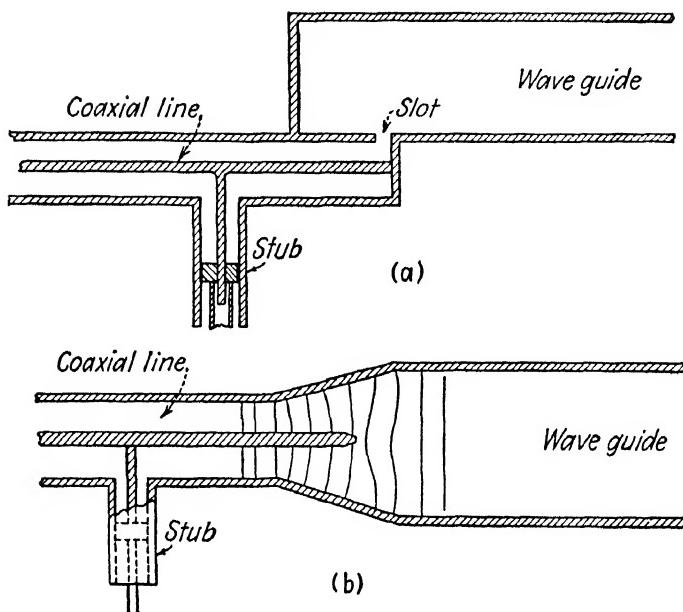
FIG. 4.—Methods of exciting wave guides from coaxial lines.
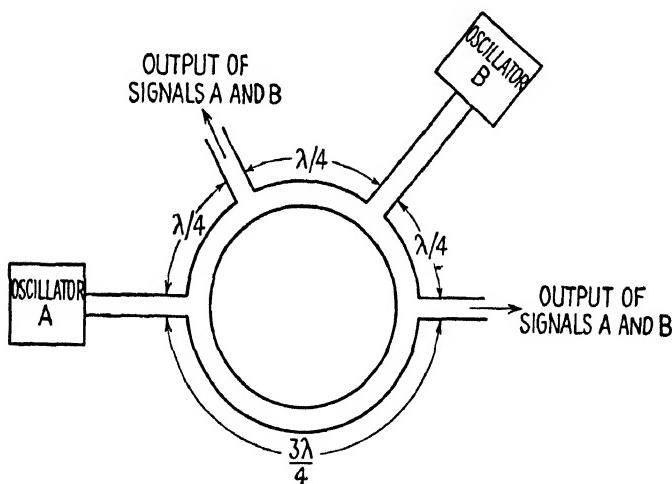
FIG. 5.—Bridge arrangement of wave guides, which prevents coupling between oscillators

guide containing several branch guides which are spaced a quarter wave-length apart on the ring. Power, entering the ring-shaped guide through any one of the branch arms, divides into two waves, one traveling in the clockwise direction and the other traveling in the counterclockwise direction around the ring. If the two waves travel the same distance to another branch arm (or distances which differ by an integral multiple of a full wave-length), they arrive at that branch arm in time phase and power will be
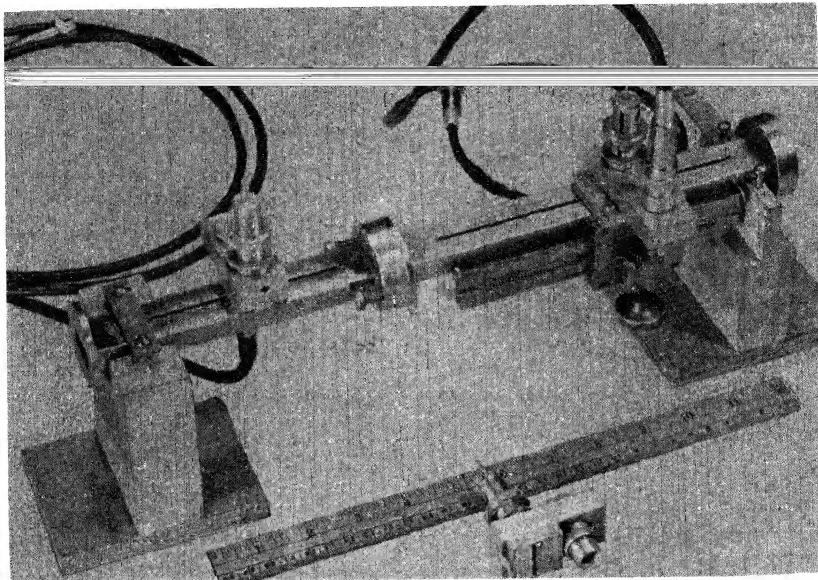
Fig. 6.—Standing-wave detector for measuring the standing-wave ratio in a wave guide
(*Courtesy of the M.I.T. Radiation Laboratory.*)

transmitted through that arm. If the two paths differ by a half wavelength, the two waves arrive in phase opposition; hence, the waves cancel and no power will be transmitted through the branch arm.

Referring to Fig. 5, it is evident, therefore, that power can be transmitted from either oscillator to the adjacent arms, but there will be no coupling between oscillators.

18.02. Impedance and Power Measurement in Wave Guides.—A measurement of the standing-wave ratio in a wave guide is useful for a number of purposes. It may be used to determine whether or not a load is properly matched to the guide and, if a mismatch occurs, how much power is sacrificed by the mismatch. The standing-wave ratio measurement may also be used to compute the effective impedance terminating a guide by a method similar to that described in Sec. 10.05. Figure 6 shows a standing-wave detector for use at wavelengths of approximately 3 centimeters. This con-

sists of a slotted section of wave guide with a movable carriage which is adjusted by a rack and pinion gear. The carriage contains a small probe antenna which protrudes through the slot into the guide. The antenna is connected to a crystal detector and thence either to a microammeter or to an amplifier with a suitable output meter. The device is then calibrated to read the d-c current. The meter may be calibrated to read the standing-wave ratio directly. By adjusting the output so that the meter has full-scale deflection with the detector in the position of maximum intensity and then moving the detector to the position of minimum intensity, the minimum deflection on the instrument gives the standing-wave ratio.
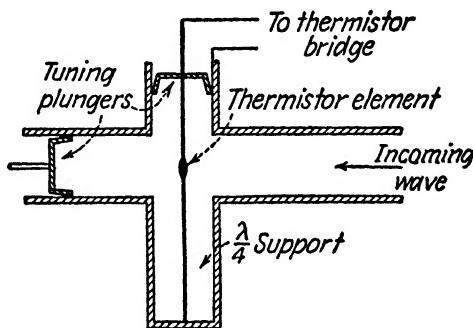
FIG. 7.—Thermistor mount in a wave guide.

The power in a wave guide may be measured by means of a thermistor or bolometer bridge, such as that described in Sec. 10.06. The thermistor or bolometer element is mounted in the guide in a position of maximum electric intensity and parallel to the electric-intensity vector, as shown in Fig. 7. The tuning plungers are used to match the element to the guide.

In many applications, it is necessary to measure the power by a "sampling" process which does not interfere with the transmission of power to the load. The device used for sampling the power must not be affected by standing waves in the guide. This may be accomplished by means of a directional coupler, such as shown in Fig. 8. The directional coupler consists of an auxiliary wave guide which is coupled to the main guide by means of two small apertures spaced a quarter wavelength apart. The auxiliary guide is terminated at one end by an absorbing medium having a wave impedance equal to the characteristic wave impedance of the guide, and at the other end by a thermistor element or a crystal detector. Since the two apertures are spaced a quarter wavelength apart in the guide, the fields at the thermistor element or crystal detector add for one direction of power flow in the main guide and cancel for the opposite direction of power flow. The directional coupler, therefore, measures power flow in one direction only and is not affected by power flow in the reverse direction. Ordi-

narily a small fraction of the total power in the guide enters the directional coupler, a typical value of the power ratio being 20 decibels. Two directional couplers, arranged to measure power flow in opposite directions, may be used to compare the power in the outgoing and reflected waves.
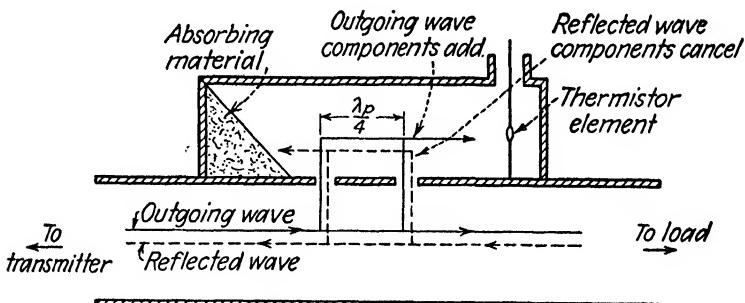
Fig. 8.—Directional coupler.

The directional coupler may also be used in reverse for launching a wave in one direction only in the guide. The thermistor is then replaced by a transmitting antenna and the coupling apertures are enlarged so as not to introduce appreciable attenuation of the signal in going from the auxiliary guide to the main guide.

18.03. The Spectrum Analyzer.—The spectrum analyzer is a useful laboratory instrument for microwave measurements. It provides a means of observing the frequencies present in a given signal, as well as the relative
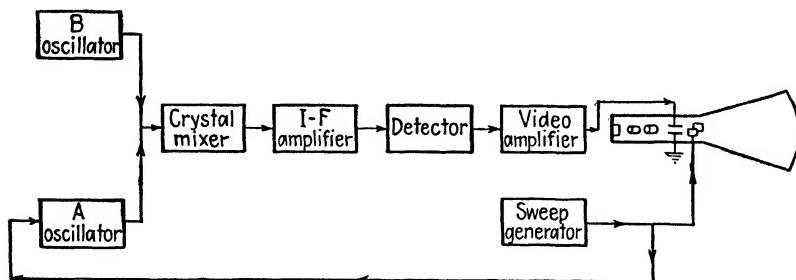
Fig. 9.—Block diagram of a spectrum analyzer.

magnitudes of the various frequencies. It can be used, for example, to measure the frequency drift of oscillators, to measure the Q of resonators, or to observe the side bands present in a modulated wave.

The block diagram of the spectrum analyzer shown in Fig. 9 contains two microwave oscillators, designated A and B. The A oscillator is frequency modulated by impressing a sawtooth voltage upon an electrode in the oscillator which is frequency-sensitive to voltage. The B oscillator operates at

a constant frequency.   The remaining components include a crystal mixer, a narrow-band intermediate-frequency amplifier, a detector, a video amplifier, and a cathode-ray oscilloscope.   The horizontal sweep of the oscilloscope is derived from the sawtooth generator which is used to frequency-modulate the A oscillator.

Assume now that the B oscillator is amplitude modulated by a pure sine-wave voltage.   The modulated wave then contains a carrier frequency and two side-band frequencies.   As the A oscillator sweeps across its frequency spectrum, the oscillator output combines at successive instants of time with the lower side band, the carrier, and the upper side band from oscillator B, to produce a difference frequency which falls within the range of the intermediate-frequency amplifier.   Therefore, the pattern on the oscilloscope contains three sharp

Fig. 10.—Spectrum of a pulsed carrier, showing the sidebands present in the wave.

vertical pulses which are separated from each other by a horizontal distance depending upon the modulation frequencies.

The horizontal axis on the oscilloscope represents the frequency scale. It can be calibrated by amplitude modulating the B oscillator with a small rectangular voltage whose frequency is accurately controlled.   The oscilloscope pattern then contains a number of equally spaced marker pulses corresponding to the frequencies $f_0 \pm nf_1$, where f_0 is the carrier frequency of the B oscillator, f_1 is the modulating frequency, and n is any integer

band frequencies can then be determined by their relative position with respect to the marker pulses.

Figure 11 shows how the spectrum analyzer may be used to measure the Q of a resonator. The resonator and a calibrated attenuator are inserted between the B oscillator and the mixer. The B oscillator is tuned to the resonant frequency of the resonator. This will be indicated by maximum height of the pulse on the oscilloscope. The calibrated attenuator is then adjusted to insert a 3-decibel loss, thereby reducing the power delivered
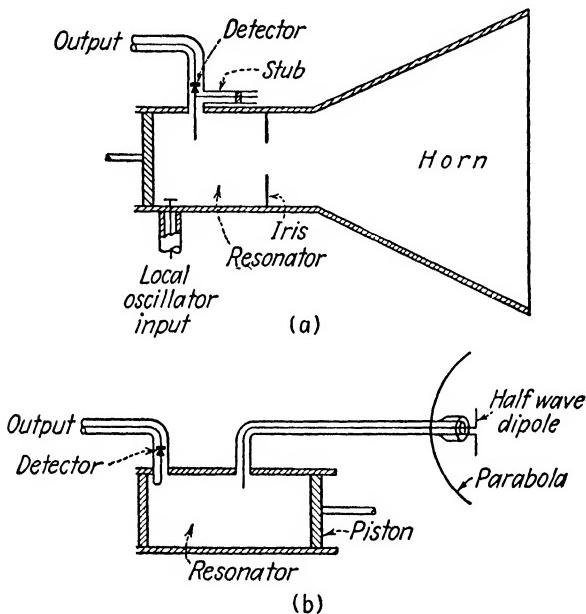
FIG. 12.—Methods of using resonators in receiving systems.

to the resonator to one-half of its former value. The height of the pulse is observed on the oscilloscope. The attenuator is then returned to its original setting and the B oscillator frequency is varied until the pulse on the oscilloscope is reduced to the previously determined half-power value. If f_0 is the resonant frequency and Δf is the change in frequency corresponding to the half-power point, the Q becomes

$$Q = \frac{f_0}{2\,\Delta f} \tag{1}$$

It is necessary that the power output of the B oscillator remain constant during this measurement, since any variation would give erroneous results.

18.04. Receiving Systems.—A wave-guide system may be either tuned or untuned. Figure 12a shows a tuned receiving system consisting of a

receiving horn which is coupled to a resonator by means of an iris aperture. A crystal detector, also coupled to the resonator, is used to detect the incoming signal. If a superheterodyne receiver is used, the local oscillator output may be coupled into the resonator by means of the probe shown in the diagram. The signal emerging from the crystal detector then contains a frequency corresponding to the difference between the carrier frequency and local oscillator frequency. The difference frequency is amplified by the
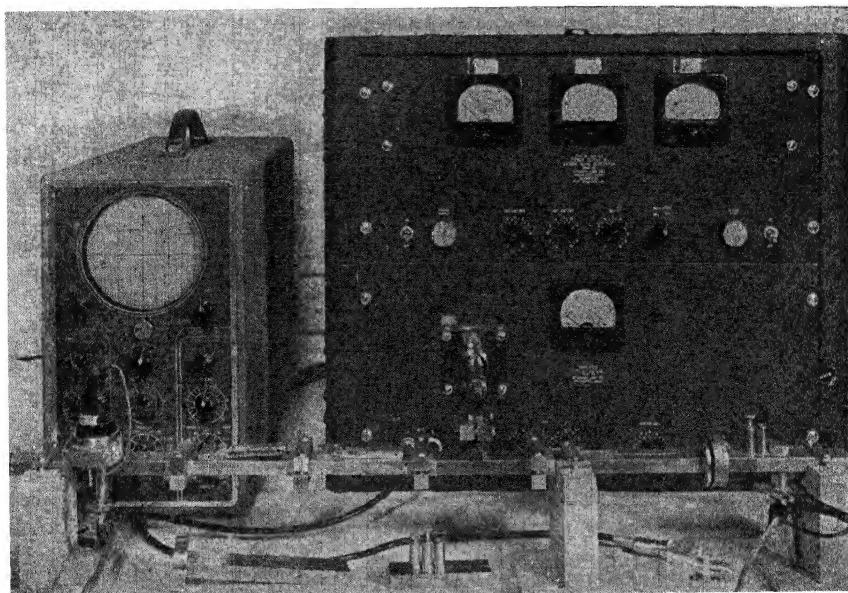
Fig. 13.—Three-centimeter laboratory setup.   (*Courtesy of the M.I.T. Radiation Laboratory.*)

intermediate-frequency amplifier following the crystal detector. A final detector then detects the signal. Precautions must be taken to use loose coupling between the local oscillator and the resonator in order to avoid frequency changes of the local oscillator due to either the tuning of the resonator or the tendency of the local oscillator to lock in with the incoming signal. In some applications it may be desirable to use separate resonators for the incoming and local-oscillator signals in order to avoid frequency-pulling of the local oscillator. The two resonators are then coupled to the detector by means of iris apertures.

A typical test-bench setup with elements which correspond to a transmitter and a receiver is shown in Fig. 13. At the extreme left is a Shepherd-Pierce tube, operating at a wavelength of 3 centimeters. The output terminal of this tube extends into the wave guide and serves as the transmitting antenna. This is followed by two "flap attenuators," each consisting of a

thin piece of fiber upon which has been deposited a layer of carbon or other absorbing material. The attenuators may be moved into or out of the wave guide to vary the attenuation. In the center of the guide is a wavemeter. This is followed by a slotted section for a standing-wave detector. The probe and detector of the standing-wave detector unit are on the table below the guide. At the end of the guide is a crystal detector, with two adjustable screws to match the impedance and thereby assure

(a)
Reflects $TE_{1,1}$ Mode

(b)
Reflects $TE_{0,m}$
Passes $TM_{0,m}$

(c)
Reflects $TM_{0,1}$

(d)
Reflects $TM_{0,1}$

(e)
Reflects $TE_{0,m}$
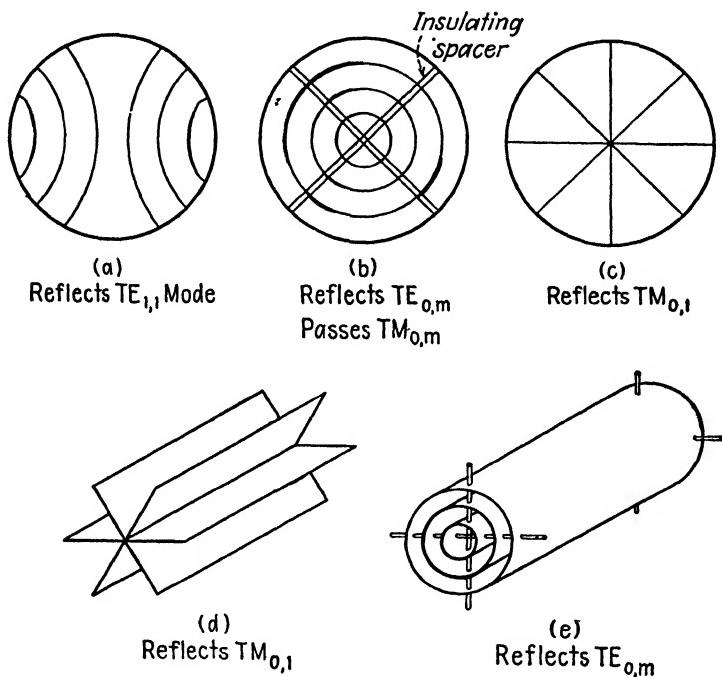
Fig. 14.—Gratings used in wave guides either to absorb or to reflect the given modes.

maximum power transfer to the detector. An adjustable piston in the end of the guide also serves to match the impedances at the receiving end of the guide.

18.05. Wire Gratings.—Wire gratings of the type shown in Fig. 14 may be used in wave guides to either reflect or absorb a particular mode without interfering with other modes. The wires are placed so as to coincide with the electric field lines for the mode which is to be reflected or absorbed. If the wires have high conductivity, the particular mode will be reflected. If the wires have low conductivity, part of the energy in the particular mode will be absorbed and part will be reflected. A second grating of similar construction, placed a quarter wavelength from the first grating, increases the power absorption. Gratings of this type are sometimes used in wave

guides and resonators to absorb the energy which appears in undesired modes. Two types of longitudinal gratings are shown in Figs. 14d and 14e. These are constructed of sheet metal and are more effective than the wire gratings.

Grating detectors, such as those shown in Fig. 15, may be used in receivers to respond to a single mode. The effectiveness of the detectors may be increased by placing an adjustable piston in the guide beyond the detectors in order to set up a standing wave in the guide which has its maximum value at the position of the detector.
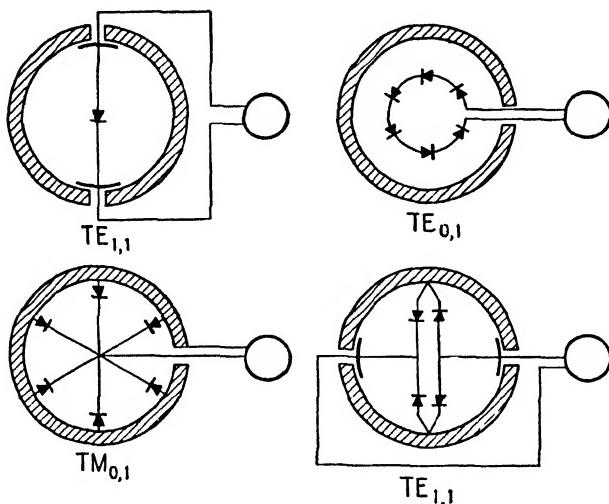
FIG. 15.— Grating detectors which respond to the modes indicated.

Wire gratings may be used to convert from one mode to another mode. Gratings for this purpose are illustrated in Fig. 16. They consist essentially of a superposition of two wire gratings corresponding to the two modes. The grating of Fig. 16a may be used to convert from a $TE_{0,1}$ mode to a $TE_{1,1}$ mode, or vice versa, while that shown in Fig. 16b converts between the $TM_{0,1}$ mode and the $TE_{0,1}$ mode. Transverse apertures, such as those shown in Figs. 16c and 16d may also be used as converters.

18.06. Multiplex Transmission through Wave Guides.—Two or more signals may be transmitted simultaneously through a wave guide and separated at the distant end. In order to separate the signals, it is necessary that they have either:

1. Different carrier frequencies but the same mode.
2. The same carrier frequency but different modes.
3. Different carrier frequencies and different modes.

If different carrier frequencies are used, the signals may be separated at the distant end either by means of resonators or by the use of a super-

$$TE_{1,1} \rightleftarrows TE_{0,1}$$

(a)

$$TM_{0,1} \rightleftarrows TE_{0,1}$$

(b)

$$TE_{1,1} \rightarrow TE_{1,1} + TM_{1,1} + TM_{2,1}$$

(c)

$$TM_{0,1} \rightarrow TM_{0,1} + TE_{2,1}$$

(d)
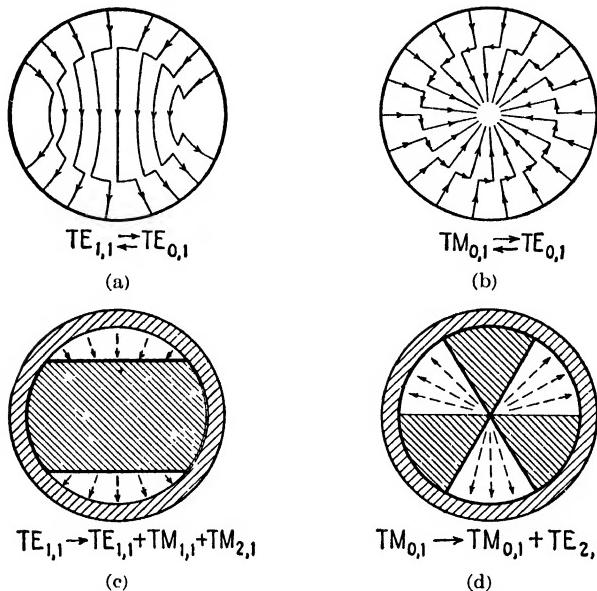
FIG. 16.—Gratings for converting from one mode to another mode in circular guides.

(a)

Detector
$TM_{0,1}$ Mode

Reflects $TM_{0,1}$
Passes $TE_{0,1}$

Detector
$TE_{0,1}$ Mode

(b)
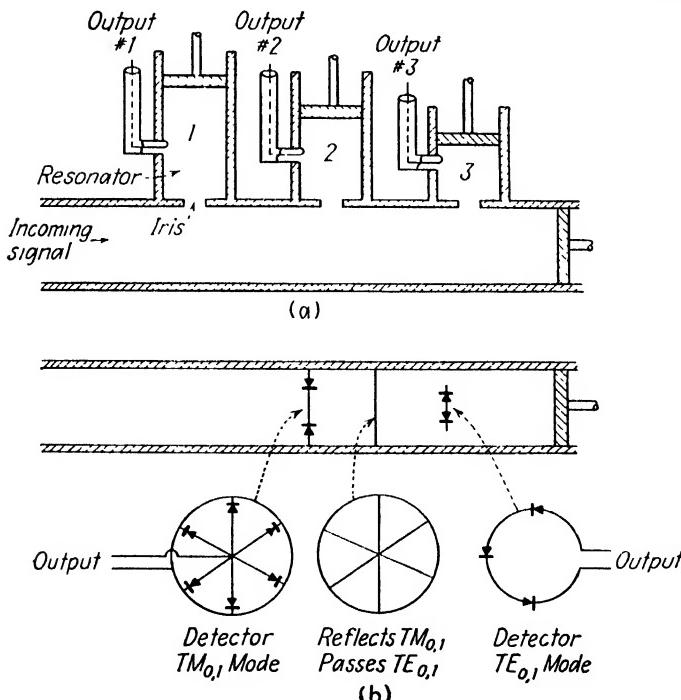
FIG. 17.—Multiplex transmission in circular guides: (a) signals having different carrier frequencies, and (b) signals transmitted in different modes.

heterodyne receiver. In the case of the superheterodyne receiver, the selectivity between signals is determined largely by the bandwidth of the intermediate-frequency amplifier. Since this amplifier can be sharply tuned, it is possible to separate signals which differ in carrier frequency by only a few megacycles.

Resonators may also be used to separate two or more signals on the basis of differences in carrier frequency. In general, however, resonators do not provide as sharp a selectivity between signals as is possible with the superheterodyne receiver. The selectivity can be greatly improved by using the resonator as a preselector which feeds into a superheterodyne receiver.
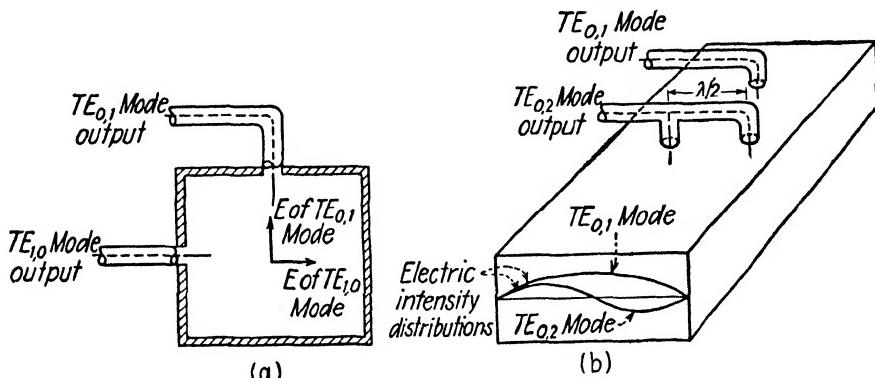
Fig. 18.—Multiplex systems employing different modes: (a) $TE_{0,1}$ modes and $TE_{1,0}$ modes, and (b) $TE_{0,1}$ and $TE_{0,2}$ modes.

Figure 17a shows a method of using resonators to separate the signals in a multiplex wave-guide system. The same arrangement may be used as a superheterodyne receiver by coupling a local oscillator to the various resonators and using the crystals for converters. The succeeding stages of amplification are then tuned to the intermediate frequency.

If the signals are to be separated on the basis of different modes, it is necessary to use mode-selective detectors, that is, detectors which respond to one mode but not to other modes. Figure 17b shows a system using a grating reflector and grating detectors. The $TM_{0,1}$ mode is reflected by the grating reflector and the corresponding detector is placed a quarter wavelength from the reflector on the generator side in order to be in a position of maximum electric intensity. The $TE_{0,1}$ mode passes through the $TM_{0,1}$ mode detector and reflector and is reflected from the end wall of the guide. Consequently, its detector is placed a quarter wavelength from the end wall. It should be noted that the quarter-wavelength distances in the guide are different for the two modes.

Other mode-selective detector systems are shown in Fig. 18. The system shown in Fig. 18a uses the $TE_{0,1}$ and $TE_{1,0}$ modes, while that of Fig. 18b

uses the $TE_{0,1}$ and $TE_{0,2}$ modes. In both systems the antennas consist of probes which are placed in a position of maximum electric intensity for the desired mode. In Fig. 18b the $TE_{0,2}$ antennas also pick up the $TE_{0,1}$ mode. However, by using a half-wavelength section of line between the two antennas, the $TE_{0,1}$ voltages can be made to cancel, leaving only the $TE_{0,2}$ mode signal.

A number of difficulties are likely to be encountered if the separation of two or more signals is attempted entirely on the basis of different modes of transmission. Any irregularities in the guide, or even the presence of probe antennas, such as those shown in Fig. 18, will tend to distort the field in the guide and introduce coupling between the various modes, resulting in objectionable cross talk.
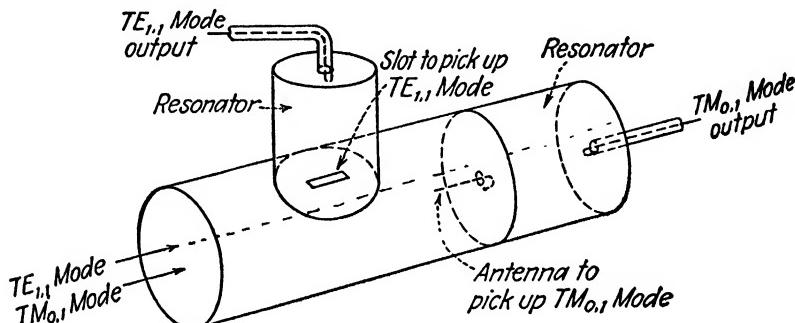
Fig. 19.—Multiplex transmission system using the $TE_{1,1}$ mode and $TM_{0,1}$ mode in a circular guide.

A more complete separation of the signals may be accomplished by combining the methods described above, that is, by using different carrier frequencies and different modes. The modes may then be separated by means of mode-selective detectors and the carrier frequencies may be separated either by resonators or by a superheterodyne receiver or both. In general, it is preferable to restrict the transmission to not more than two modes, since it becomes difficult to obtain suitable mode-selective detectors for a larger number of modes. It should be noted that the use of different modes makes it possible not only to separate the signals at the receiving end, but also to isolate the transmitters at the sending end. This serves to minimize the possibility of interaction of the transmitters.

Figure 19 shows a method of separating the two modes which have the longest wavelength in a circular guide, *i.e.*, the $TE_{1,1}$ mode and the $TM_{0,1}$ mode. If the two signals are also transmitted on different carrier frequencies, resonators may be used to improve the selectivity between the signals. A slot in the guide wall is used to intercept the $TE_{1,1}$ mode. In general, a slot will intercept a given mode only if it interrupts the current flow in the wall of the guide for that mode. Since the $TE_{1,1}$ mode has

currents flowing in the ϕ direction in the guide wall, a longitudinal slot will interrupt these currents; hence, this mode will be transmitted through the slot. The $TM_{0,1}$ mode currents flow only in the longitudinal direction; hence, the slot will have very little effect upon this mode.

18.07. Wave-guide Filters.—Several wave-guide filters are shown in Fig. 20. These are band-pass filters with pass bands corresponding to harmonics of the first pass band. In the filter of Fig. 20a, the centers of the pass bands
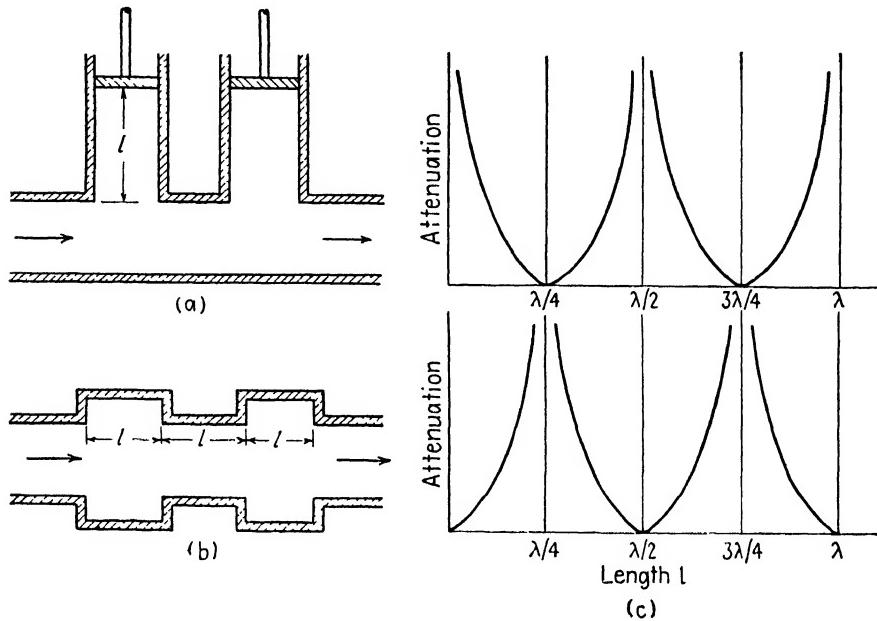
FIG. 20.—Filters using wave-guide sections.

occur at the frequencies for which the short-circuited sections of wave guide have a maximum input wave impedance. Maximum attenuation occurs when the input wave impedance of the short-circuited sections is a minimum.

The filter shown in Fig. 20b has the center of the pass band at the frequency for which the enlarged section of the guide is an integral number of half wavelengths long. A half-wavelength section of guide is similar to a half-wavelength transmission line in that it serves as a one-to-one ratio transformer, that is, the input wave impedance is equal to the terminal impedance. Hence, there is no apparent discontinuity in impedance at the wavelength for which the enlarged sections are a half wavelength long, whereas, at other frequencies there is an abrupt discontinuity at the junctions, resulting in reflections. Maximum attenuation occurs at the frequencies for which the enlarged sections are a quarter wavelength long.

CHAPTER 19

LINEAR ANTENNAS AND ARRAYS

The function of an antenna is either to radiate or to receive electromagnetic energy. A transmitting antenna has an alternating emf applied to its terminals which produces a current in the antenna and an electromagnetic field in space. The energy radiated from the antenna appears as a traveling wave, propagating outward from the antenna with a velocity equal to the velocity of light. A receiving antenna, placed in an electromagnetic field, has an emf induced in it by the field which produces an alternating potential difference at the antenna terminals.

A given antenna or an array of antennas has similar characteristics when used either as a transmitting or as a receiving antenna. For example, the directional properties of an antenna system are the same when the antenna is used as a transmitting antenna as when used as a receiving antenna, under similar conditions of operation. This is a consequence of an important principle of reciprocity which will be discussed in the following chapter. We shall consider the antenna primarily from the viewpoint of the transmitting antenna, although many of the conclusions apply equally well to receiving antennas.

Two or more antennas may be grouped in an *array* to obtain directional radiation. The directional characteristics of the array are determined by the spacing between antennas and the phase relationships of the currents in the various antennas of the array. Since antennas used at microwave frequencies have small physical size, they are particularly adaptable for use in arrays, parabolic reflectors, and other directional radiating systems. This makes it possible to concentrate the radiated energy into a narrow beam for point-to-point communication, thereby affecting a very appreciable saving in transmitter power.

In the design of an antenna system, the following factors must be considered:

1. Design of the antenna system so as to obtain the desired field distribution in space.

2. Determination of the total radiated power and radiation resistance.

3. Determination of the input impedance as a function of frequency. This is particularly important when considering wide-band antennas.

4. Design of the electrical networks which feed the antenna system. These networks must be such that the antenna presents the proper impedance to the transmitter. If maximum power output is desired, the imped-

ances must be matched at the transmitter. For antenna arrays, the networks must be adjusted to give the proper magnitude and phase of currents in the various antennas of the array.

This chapter will deal with the methods of determining the field distribution of linear antennas and arrays. The Poynting-vector method of evaluating the total radiated power and the radiation resistance of single antennas will be described. The determination of self-impedances and mutual impedances of antennas requires a more accurate determination of the field distribution in the vicinity of the antenna than that presented in this chapter. Consequently, this subject will be deferred for treatment in the following chapter.

19.01. Methods of Determining the Field Distribution of an Antenna.— In the foregoing chapters we have considered solutions of Maxwell's equations as applied to passive systems of relatively simple geometry. We started with the general solution of the wave equation in the coordinate system which was best suited to the boundaries of the particular problem. The constants in the general solution were then adjusted so as to satisfy the boundary conditions. No attempt was made to relate the fields to the charges or currents at the source, but rather, it was assumed *per se* that a proper distribution of charges and currents could exist which would produce the given field. This type of solution reveals all possible types of modes which can exist within a given set of physical boundaries, but it does not specify which modes actually do exist for a given distribution of charges and currents at the source.

We found that in wave guides and resonators having relatively simple geometry the boundary conditions favored certain modes and discouraged others. Thus, a wave guide having dimensions such as to pass only the dominant mode will transmit the dominant mode but attenuate all higher modes. Consequently, the physical boundaries may be such as to favor certain modes, the existence of which, however, is contingent upon a proper distribution of charges and currents at the source.

We could presumably determine the field distribution of an antenna in much the same manner as that used for wave guides and resonators. That is, starting with the general solution of the wave equation, we would proceed to evaluate the constants in this equation in such a manner as to satisfy the boundary conditions. If we assumed perfectly conducting boundaries, the summation of the tangential components of electric intensity for the various modes would have to be zero at the conducting surfaces. Similarly, the summation of the tangential components of magnetic intensity would be equal to the surface current density. If we were to pursue this course, we would find that a very large number of modes would be required to completely describe the field of an antenna and the problem of evaluating the constants so as to satisfy the boundary conditions

would involve serious mathematical difficulties. A few types of antennas having relatively simple geometry, such as the sphere, the spheroid, and the biconical antenna, have been analyzed by this method,[1,2,3] but the analysis is too involved for ordinary engineering work.

Fortunately, there are more direct ways of determining the field distribution of an antenna which involve less mathematical complication. The method which will be used here starts with the vector potential as given by Eq. (15.02-12),

$$\bar{A} = \frac{\mu}{4\pi} \int_\tau \frac{J_c[t - (r/v_c)]}{r}\, d\tau \qquad (15.02\text{-}12)$$

If the current distribution in the antenna is known, Eq. (15.02-12) may be used to evaluate the vector potential anywhere in space. The magnetic intensity is then obtained by applying

$$\bar{H} = \frac{1}{\mu}\, \nabla \times \bar{A} \qquad (15.02\text{-}1)$$

and the electric intensity (for sinusoidally varying fields in a lossless medium) is obtained from

$$\nabla \times \bar{H} = j\omega\epsilon\bar{E} \qquad (13.06\text{-}2)$$

In applying this method, it is necessary to start with a known current distribution in the antenna. The simplest case is that of a linear antenna, since it can be shown that the current distribution in an infinitely thin straight wire antenna has a sinusoidal variation along the length of the antenna.[4] We shall first consider the field of an incremental antenna which is assumed to have infinitesimal length. The linear antenna will then be treated as consisting of a large number of infinitesimal antennas placed end to end.

19.02. Field of an Incremental Antenna.—Consider the incremental antenna shown in Fig. 1, having a length dz and carrying a uniform current $Ie^{j\omega t}$.

Referring to the expression for the vector potential, Eq. (15.02-14), we replace $J_c d\tau$ by $I\, dz$, yielding the vector potential at a point distant r from the antenna,

$$A_z = \frac{\mu I\, dz}{4\pi r}\, e^{j(\omega t - \beta r)} \qquad (1)$$

[1] STRATTON, J. A., "Electromagnetic Theory," pp. 554–560, McGraw-Hill Book Company, Inc., New York, 1941.

[2] CHU, L. J., and J. A. STRATTON, Forced Oscillations of Prolate Spheroid, *J. Applied Phys.*, vol. 12, pp. 241–248; March, 1941.

[3] SCHELKUNOFF, S. A., "Electromagnetic Waves," pp. 441 ff., D. Van Nostrand Company, Inc., New York, 1943.

[4] SCHELKUNOFF, S. A., "Electromagnetic Waves," pp. 142–143, D. Van Nostrand Company, Inc., New York, 1943.

Expressing the vector potential in spherical coordinates and dropping the time function $e^{j\omega t}$, we obtain

$$A_r = A_z \cos\theta = \frac{\mu I\,dz}{4\pi r}\cos\theta e^{-j\beta r} \tag{2}$$

$$A_\theta = -A_z \sin\theta = -\frac{\mu I\,dz}{4\pi r}\sin\theta e^{-j\beta r} \tag{3}$$

$$A_\phi = 0$$
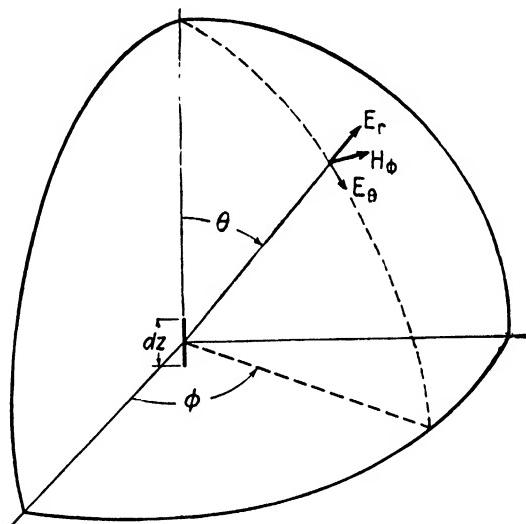
FIG. 1.—Coordinates for the incremental antenna.

The magnetic intensity is obtained by inserting A_r and A_θ into $\bar{H} = (1/\mu)\nabla \times \bar{A}$, where $\nabla \times \bar{A}$ is given in Appendix III. With the additional substitution of $\beta = 2\pi/\lambda$ and remembering that $\partial/\partial\phi = 0$, we obtain

$$H_r = H_\theta = 0$$

$$H_\phi = \left(\frac{j2\pi}{\lambda r} + \frac{1}{r^2}\right)\frac{I\,dz}{4\pi}\sin\theta e^{-j\beta r} \tag{4}$$

To obtain the electric intensity, insert H_ϕ from Eq. (4) into $\nabla \times \bar{H} = j\omega\epsilon\bar{E}$, giving

$$E_r = \eta\left(\frac{1}{r^2} - \frac{j\lambda}{2\pi r^3}\right)\frac{I\,dz}{2\pi}\cos\theta e^{-j\beta r} \tag{5}$$

$$E_\theta = \eta\left(\frac{j2\pi}{\lambda r} + \frac{1}{r^2} - \frac{j\lambda}{2\pi r^3}\right)\frac{I\,dz}{4\pi}\sin\theta e^{-j\beta r} \tag{6}$$

The electric and magnetic intensities contain terms varying as $1/r$, $1/r^2$, and $1/r^3$. The components containing $1/r^2$ and $1/r^3$ predominate in the immediate vicinity of the antenna and are known as the *induction field* of the antenna. The induction field represents reactive energy which is stored in the field during one portion of the cycle and returned to the source during a later portion of the cycle. The induction field terms become vanishingly small at remote distances from the antenna and hence do not contribute to the radiation of power from the antenna.

The terms varying as $1/r$ in the intensity expressions comprise the *radiation field* of the antenna. The radiation field is comprised of electromagnetic waves traveling radially outward with a propagation factor $e^{j(\omega t - \beta r)} = e^{j\omega(t - r/v_c)}$ and with intensities which vary inversely as the first power of the distance from the source. Equation (4) shows that the induction and radiation field components of magnetic intensity are equal at a distance of $r = \lambda/2\pi$, or approximately a sixth of a wavelength from the antenna.

It is interesting to observe that if we had assumed that the field builds up instantaneously throughout space, *i.e.*, if we had used $e^{j\omega t}$ instead of $e^{j\omega t - \beta r}$ in Eq. (1), the radiation-field terms would not have been present in the resulting intensity equations. Radiation is therefore dependent upon the fact that the field has a finite velocity of propagation. In conventional circuit analysis it is customary to ignore the finite velocity of propagation of the field. This approximation is valid if most of the field is confined to a region which is very small in comparison with the wavelength. It leads to what is known as the *quasi-stationary analysis*.

The radiation-field terms, taken alone, comprise a spherical *TEM* wave propagating radially outward from the source with a wave impedance equal to the intrinsic impedance of free space. Discarding the j factor in Eqs. (4) and (6), we obtain the radiation field intensities,

$$H_\phi = \frac{I\,dz}{2\lambda r} \sin\theta e^{-j\beta r} \qquad (7)$$

$$E_\theta = \frac{\eta I\,dz}{2\lambda r} \sin\theta e^{-j\beta r} \qquad (8)$$

The ratio of electric to magnetic intensity is equal to the intrinsic impedance of the medium, thus

$$\frac{E_\theta}{H_\phi} = \eta \qquad (9)$$

The power radiated by the incremental antenna is found by integrating the normal component of Poynting's vector over the surface of an imaginary sphere having the antenna at its center. For convenience we choose a

sphere which is large enough so that the induction-field terms are negligible. Inserting E_θ from Eq. (8) into (14.04-3) and dropping the phase-shift term $e^{-j\beta r}$, we obtain for the time-average power density,

$$p = \frac{|E_\theta|^2}{2\eta} = \frac{\eta(I\,dz)^2}{8r^2\lambda^2}\sin^2\theta \tag{10}$$
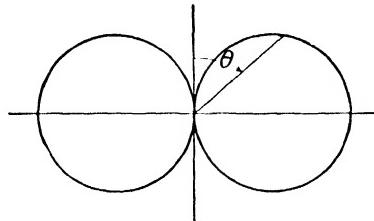
FIG. 2.—Field pattern of the incremental antenna in the vertical plane.

The total radiated power is therefore

$$P = \int_0^\pi p\,2\pi r^2 \sin\theta\,d\theta$$

$$= \frac{\eta(I\,dz)^2}{8r^2\lambda^2}\int_0^\pi 2\pi r^2 \sin^3\theta\,d\theta$$

$$= \frac{\pi\eta}{3\lambda^2}(I\,dz)^2 \tag{11}$$

The radiated power is independent of the radius of the sphere over which the power density is integrated. This is a consequence of assuming a lossless transmission medium.

The radiation resistance R_0 is defined as the ohmic resistance which would consume the same power as the antenna radiates, if the resistance were carrying the same current. The radiation resistance of the incremental antenna is

$$R_0 = \frac{2P}{I^2} = \frac{2\pi\eta}{3}\left(\frac{dz}{\lambda}\right)^2 \tag{12}$$

The field pattern of an antenna is a polar plot of the electric intensity in a given plane, taken at a constant distance from the antenna. The incremental antenna radiates uniformly in the horizontal direction; hence its field pattern in a horizontal plane is a circle. In the vertical plane the intensities vary as $\sin\theta$ and the field pattern is as shown in Fig. 2, with maximum intensity in a plane perpendicular to the antenna and zero intensity at any point directly above or below the antenna.

$t=\frac{1}{4}T$ $t=\frac{3}{8}T$

$t=\frac{1}{2}T$ $t=\frac{5}{8}T$

$t=\frac{3}{4}T$ $t=\frac{7}{8}T$
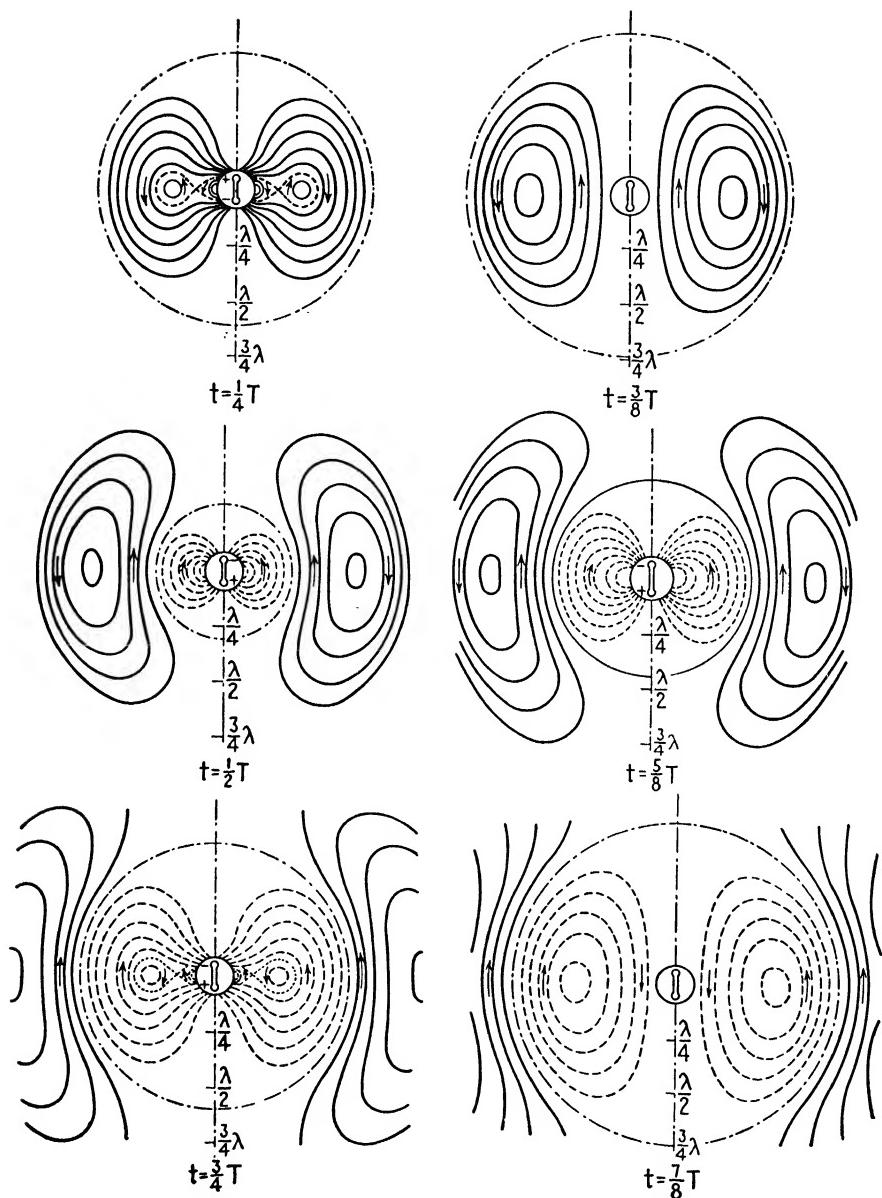
Fig. 3.—Electric field of a radiating dipole during the formation of a wave.

Figure 3 shows the electric-field lines for an oscillating dipole at various stages during the formation of a wave. The dipole considered here consists of two opposite charges which are separated an infinitesimal distance apart and which have oscillating magnitudes. Such a dipole can be shown to be

equivalent to an antenna of infinitesimal length. At the outset, the electric lines are attached to the charges, but as the charges reverse their positions, we can visualize the electric-field lines as becoming detached and forming closed loops. The magnetic-field lines are circles which are concentric with the axis of the dipole. As the wave propagates outward from the source, the closed loops of electric and magnetic intensity expand. At a remote distance from the antenna, the wave is essentially a spherical *TEM* wave, traveling radially outward with a velocity equal to the velocity of light.

19.03. Radiation Field of a Linear Antenna—Approximate Method.—A linear antenna may be viewed as consisting of a very large number of incremental antennas placed end to end. The field of the linear antenna may therefore be obtained by integrating the contributions of all of the incremental antennas.

FIG. 4.—Linear antenna.

Consider the linear antenna of Fig. 4, which is assumed to be isolated in space and to have a length l. The current in an infinitely thin antenna has a sinusoidal distribution; hence we let

$$I = I_0 \sin (\beta z + \varphi) \tag{1}$$

To obtain an expression for the electric intensity of the linear antenna, insert I from Eq. (1) into (19.02-8) and integrate. This gives

$$E_\theta = \frac{\eta \sin \theta}{2\lambda} \int_{-l/2}^{l/2} \frac{I_0 \sin (\beta z + \varphi)}{r} e^{-j\beta r} \, dz \tag{2}$$

For distances remote from the antenna, we may substitute r_0 for r in the denominator of Eq. (2). However, the term $e^{-j\beta r}$ determines the phase of the electric intensity and must be evaluated more accurately. In this term we substitute $r = r_0 - z \cos \theta$ and Eq. (2) then becomes

$$E_\theta = \frac{\eta I_0 \sin \theta}{2\lambda r_0} e^{-j\beta r_0} \int_{-l/2}^{l/2} \sin (\beta z + \varphi) \epsilon^{j\beta z \cos \theta} \, dz \tag{3}$$

Assume now that the antenna is an integral number of half wavelengths long, *i.e.*, $l = n\lambda/2$, where n is any positive integer. If the antenna is an odd integral number of half wavelengths long, the current distribution is a cosine function of z, hence we let $\varphi = \pi/2$. For an even integral number
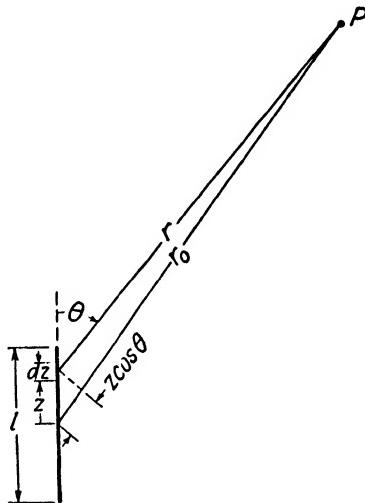
of half wavelengths long, the current distribution is sinusoidal and we let $\varphi = 0$. Integration of Eq. (3) for the two cases yields

$$E_\theta = \frac{\eta I_0}{2\pi r_0} \frac{\cos\left[(n\pi/2)\cos\theta\right]}{\sin\theta} e^{-j\beta r_0} \qquad n \text{ is odd} \qquad (4)$$

$$E_\theta = \frac{\eta I_0}{2\pi r_0} \frac{\sin\left[(n\pi/2)\cos\theta\right]}{\sin\theta} e^{-j\beta r_0} \qquad n \text{ is even} \qquad (5)$$

(a) n=1

(b) n=2

(c) n=4

(d) n=16
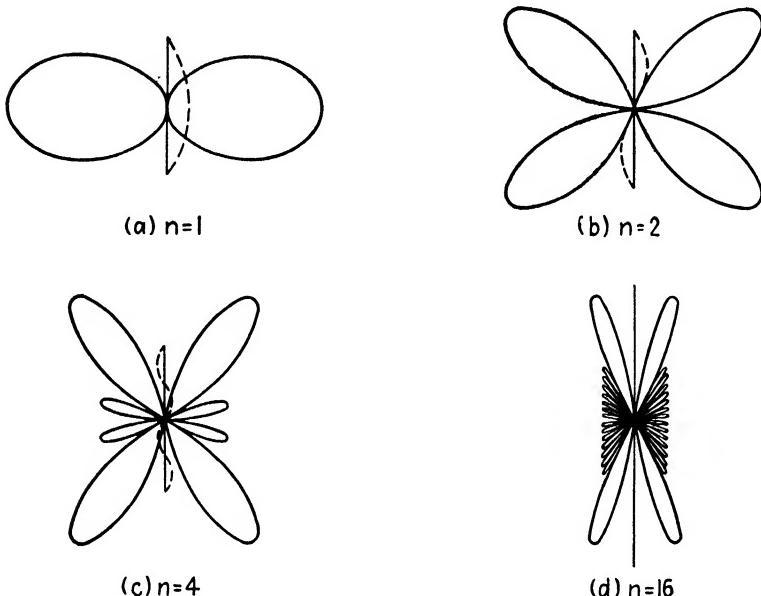
FIG. 5.—Field patterns in the vertical plane of antennas having various lengths.

Evaluating the coefficients and remembering that $e^{-j\beta r_0}$ has a magnitude of unity, we obtain the intensity

$$E_\theta = \frac{60I_0}{r_0} \frac{\cos\left[(n\pi/2)\cos\theta\right]}{\sin\theta} \qquad n \text{ is odd} \qquad (6)$$

$$E_\theta = \frac{60I_0}{r_0} \frac{\sin\left[(n\pi/2)\cos\theta\right]}{\sin\theta} \qquad n \text{ is even} \qquad (7)$$

For a half-wavelength dipole antenna, the electric intensity is

$$E_\theta = \frac{60I_0}{r_0} \frac{\cos\left[(\pi/2)\cos\theta\right]}{\sin\theta} \qquad (8)$$

The radiation field patterns in the vertical plane for antennas of various lengths are shown in Fig. 5. As the antenna length increases, the number of lobes increases and the angle $\theta_{\max}$, corresponding to the major lobe,

decreases. The number of lobes is $2n$. If the conductor were infinitely long and perfectly conducting, all of the energy would propagate in the direction of the wire.

To obtain the time-average radiated power, insert Eq. (6) or (7) into (14.04-3) and integrate over the surface of a sphere, and we have

$$P = \int_0^\pi \frac{|E_\theta^2|}{2\eta} 2\pi r^2 \sin\theta \, d\theta$$

$$= 30I_0^2 \int_0^\pi \frac{\cos^2\left[(n\pi/2)\cos\theta\right]}{\sin\theta} \, d\theta \qquad n \text{ is odd} \qquad (9)$$

$$P = 30I_0^2 \int_0^\pi \frac{\sin^2\left[(n\pi/2)\cos\theta\right]}{\sin\theta} \, d\theta \qquad n \text{ is even} \qquad (10)$$

Integration of these equations yields [1]

$$P = 15I_0^2[2.415 + \ln n - Ci2\pi n] \qquad n \text{ odd or even} \qquad (11)$$

where

$$Cix = -\int_x^\infty \frac{\cos x}{x} \, dx \qquad (12)$$

is the *cosine integral*. Values of Cix as a function of x may be obtained from tables.[2] Figure 6 shows a plot of the cosine integral as a function of x. The *sine integral*, defined by

$$Six = \int_0^x \frac{\sin x}{x} \, dx \qquad (13)$$

is also shown in Fig. 6. The sine integral will be used later in connection with the evaluation of antenna impedances.

The radiation resistance for the linear antenna is

$$R_0 = \frac{2P}{I_0^2}$$

$$= 72.45 + 30 \ln n - 30Ci2\pi n \qquad n \text{ is odd or even} \qquad (14)$$

For a half-wavelength antenna we have $n = 1$, $P = 36.56I_0^2$, and $R_0 = 73.13$ ohms. Figure 7 shows how the radiation resistance of antennas varies with the length of the antenna. The radiation resistance is the resistive component of the input impedance if the antenna is fed at the current loop (maximum current point).

---

[1] STRATTON, J. A., "Electromagnetic Theory," pp. 438–444, McGraw-Hill Book Company, Inc., New York, 1941. These equations may also be integrated graphically for specific cases.

[2] JAHNKE, E., and F. EMDE, "Tables of Functions," pp. 6–9, Dover Publications, 1943.
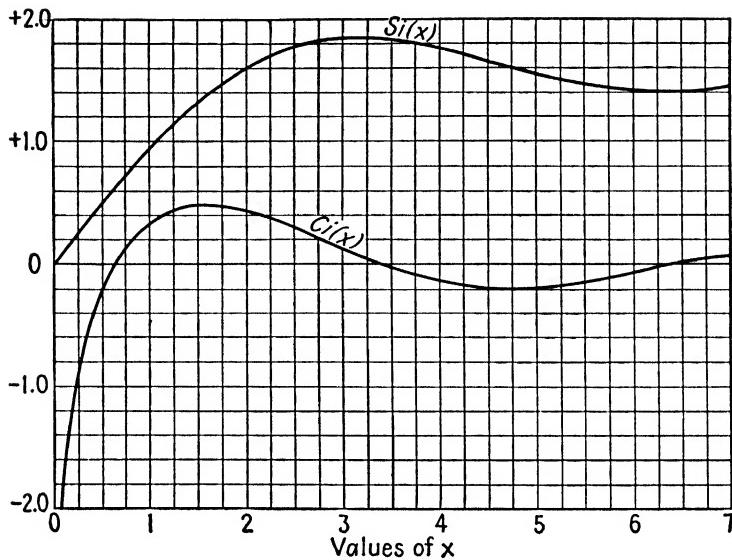
Fig. 6.—Plot of the cosine and sine integrals.
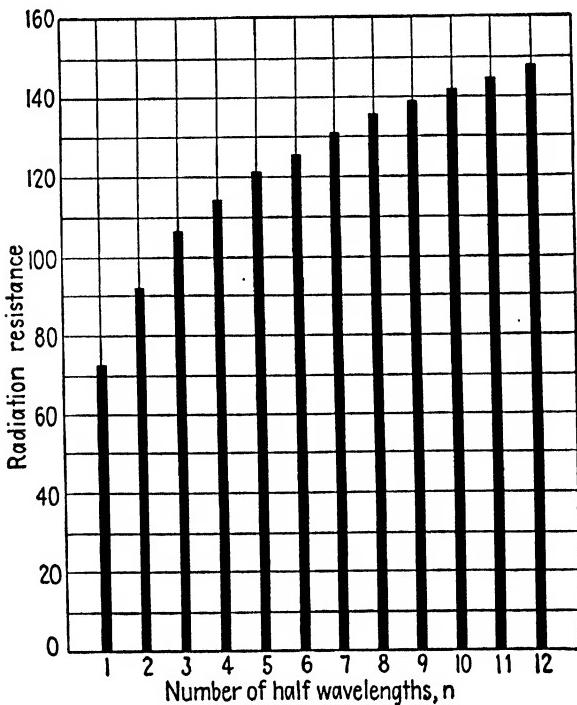
Fig. 7.—Radiation resistance of antennas of various lengths.

19.04. Antennas in the Vicinity of a Conducting Plane.—Thus far we have considered only the idealized case of an antenna which is isolated in space. If an antenna is in the vicinity of the earth or other reflection objects, the radiating characteristics of the antenna may be appreciably altered. For example, consider a vertical antenna above a perfectly conducting plane, as shown in Fig. 8a. Here we may use the principle of images and replace the antenna of Fig. 8a by the antenna and its image as
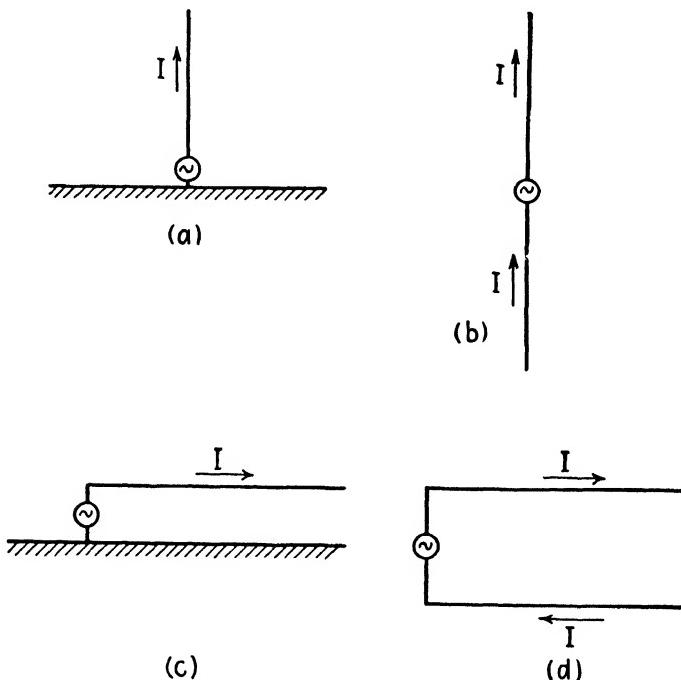
FIG. 8.—Antennas above a perfectly conducting plane.

shown in Fig. 8b. But the antenna of Fig. 8b is merely the linear antenna for which we have already derived the expressions for the field intensities. Since there is no field below the conducting plane, the radiated power is integrated over the surface of a hemisphere instead of a sphere. Consequently the radiated power and radiation resistance are one-half of the values given by Eqs. (19.03-11 and 14). Thus, the radiation resistance of a quarter-wavelength vertical antenna above a perfectly conducting ground plane is 36.56 ohms.

The horizontal antenna above a perfectly conducting ground plane, shown in Fig. 8c, may be replaced by the image equivalent of Fig. 8d. The antenna of Fig. 8d is essentially a two-element array which may be analyzed by the methods of the following section.

The electrical characteristics of the earth vary in different localities, depending upon the soil composition and moisture content. In many cases the error is not serious if perfect conductivity is assumed for the purpose of image calculations.

19.05. Radiation Field of Arrays of Linear Antenna Elements.—By combining two or more linear antennas in an array, with proper spacing between antennas and proper phasing of antenna currents, directional radiation may be obtained. Consider the array shown in Fig. 9, composed of m parallel linear antennas, each a half wavelength long, with a separation distance d between successive antennas. It is assumed that the current in each successive antenna lags the current in the preceding antenna by a phase angle of α radians.

The electric intensity at a distant point P in a horizontal plane perpendicular to the antennas, due to antenna 1, is obtained by setting $\theta = \pi/2$ in Eq. (19.03-6), which gives

$$E_{\theta 1} = \frac{60 I_0}{r_0} \qquad (1)$$

(a)-PLAN VIEW

(b)-FRONT VIEW
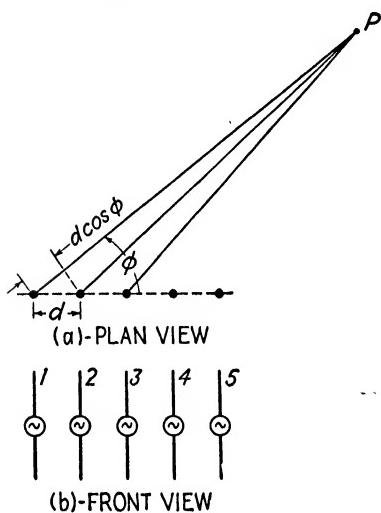
FIG. 9.—Array of linear antennas.

Antenna 2 is closer to the point P than antenna 1 by an amount $d \cos \phi$. Consequently the intensity at P due to antenna 2 leads that of antenna 1 by the angle $(2\pi d/\lambda) \cos \phi - \alpha$, where α is the phase angle between the antenna currents. Similarly, the intensity due to antenna 3 leads that of antenna 1 by an angle $2[(2\pi d/\lambda) \cos \phi - \alpha]$. The intensities at P due to the various antennas are therefore

$$E_{\theta 2} = \frac{60 I_0}{r_0} e^{j[(2\pi d/\lambda)\cos\phi - \alpha]}$$

$$E_{\theta 3} = \frac{60 I_0}{r_0} e^{j2[(2\pi d/\lambda)\cos\phi - \alpha]} \qquad (2)$$

$$E_{\theta 4} = \frac{60 I_0}{r_0} e^{j3[(2\pi d/\lambda)\cos\phi - \alpha]}$$

To simplify the notation, let

$$\delta = \frac{2\pi d}{\lambda} \cos \phi - \alpha \qquad (3)$$

The resultant intensity is then

$$E_\theta = \frac{60I_0}{r_0}[1 + e^{j\delta} + e^{j2\delta} + \cdots e^{j(m-1)\delta}]$$

$$= \frac{60I_0}{r_0}\frac{\sin(m\delta/2)}{\sin(\delta/2)} \tag{4}$$

The series form of Eq. (4) shows that maximum intensity occurs when all of the intensities are in time phase at P, that is, when $\delta = 0, 2\pi$, etc. Let us assume that $\delta = 0$. Equation (3) then gives

$$\cos\phi_{max} = \frac{\alpha\lambda}{2\pi d} \tag{5}$$

where ϕ_{max} is the value of ϕ corresponding to maximum radiation. The intensity in this direction is $E_\theta = 60I_0 m/r_0$, this being m times the intensity of a single antenna with the same current (but not necessarily the same power). By a proper choice of the antenna spacing d and phase angle α, any desired value of ϕ_{max} may be obtained.

The summed form of Eq. (4) shows that if m is large, secondary lobe maxima occur approximately when $m\delta/2 = k\pi/2$, where k is an odd integer. Nodal values occur when k in this expression is an even integer, provided that $\sin(\delta/2)$ is not also zero. The latter case results in an indeterminant value of Eq. (4). The corresponding values of $\cos\phi_{max}$ and $\cos\phi_0$ are found by substituting the values of δ from this expression into Eq. (3), yielding

$$\cos\phi_{max} = \left(\frac{k\pi + m\alpha}{2\pi md}\right)\lambda \qquad k \text{ is odd} \tag{6}$$

$$\cos\phi_0 = \left(\frac{k\pi + m\alpha}{2\pi md}\right)\lambda \qquad k \text{ is even} \tag{7}$$

Two particular cases of interest are the *broadside array*, with maximum radiation corresponding to $\phi_{max} = \pi/2$, and the *end fire* array, with $\phi_{max} = 0$ or π radians. Let us first consider the broadside array.

In the broadside array, we have $\cos\phi_{max} = 0$, and Eq. (5) gives

$$\frac{\alpha\lambda}{2\pi d} = 0 \tag{8}$$

This is satisfied by $\alpha = 0$, requiring that all antennas be fed in the same phase regardless of the spacing between antennas. A single-row broadside antenna has two major lobes which are 180 degrees apart as shown in Fig. 10. The angle ϕ corresponding to secondary maxima and nodes, for

arrays having a large number of elements, may be found by setting $\alpha = 0$ in Eqs. (6) and (7), giving

$$\cos \phi_{\max} = \frac{k\lambda}{2md} \qquad k \text{ is odd} \qquad (9)$$

$$\cos \phi_0 = \frac{k\lambda}{2md} \qquad k \text{ is even} \qquad (10)$$

The directivity of the broadside array increases as the over-all length of the array increases. The directivity improves somewhat as the spacing between

(a) $l = 2\lambda$ (b) $l = 4\lambda$ (c) $l = 8\lambda$
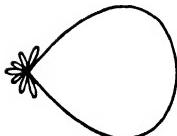
(d) $l = 2\lambda$ (e) $l = 4\lambda$ (f) $l = 8\lambda$

Fig. 10.—Field patterns of arrays of various lengths. Upper row, broadside antennas; lower row, end-fire antennas.

antenna elements is increased up to a critical value of approximately $d = 3\lambda/4$, beyond which the directivity rapidly decreases.

The end-fire array has a single major lobe with maximum radiation at $\phi_{\max} = 0$ or π radians. Since $\cos \phi_{\max} = \pm 1$, Eq. (5) becomes

$$\alpha = \pm \frac{2\pi d}{\lambda} \qquad (11)$$

Inserting this value of α into Eqs. (6) and (7), we could obtain the values of ϕ corresponding to the secondary lobe maxima and the nodes.

The end-fire array has a single major lobe with maximum radiation in the direction of the end of the array having the lagging phase. The major

lobe width, however, is wider than that of the broadside antenna for a given array. The directivity increases with the over-all length of the array and also increases as the spacing between conductors increases up to a critical value of approximately $d = 3\lambda/8$.
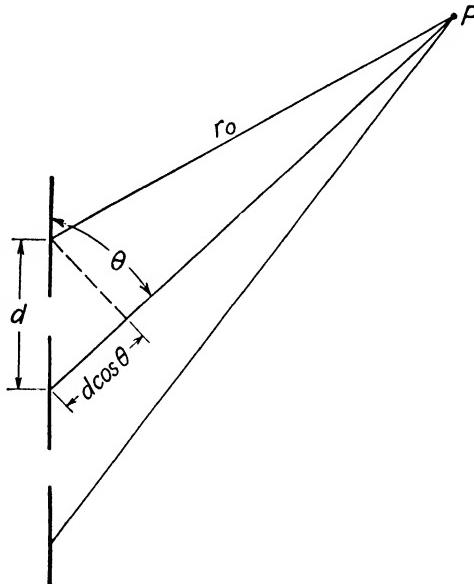
FIG. 11.—Colinear antenna array.

19.06. Other Types of Arrays.—The colinear antenna array, shown in Fig. 11, may be treated by methods similar to those of the preceding article. The electric intensity at a distant point P is

$$E_\theta = \frac{60I_0}{r_0}[1 + e^{-j[(2\pi d/\lambda)\cos\theta - \alpha]} + e^{-j2[(2\pi d/\lambda)\cos\theta - \alpha]} + \cdots]e^{-j\beta r_0}f(\theta)$$

$$= \frac{60I_0}{r_0}e^{-j\beta r_0}\frac{\sin(m\delta/2)}{\sin(\delta/2)}f(\theta) \tag{1}$$

where m is the number of antenna elements and δ is given by Eq. (19.05-3) with ϕ replaced by θ. The function $f(\theta)$ is given in Eqs. (19.03-6 and 7). This array radiates uniformly in the ϕ direction but may be made to have any desired directivity in the θ direction.

Arrangements for feeding the antenna elements of colinear arrays are shown in Fig. 12. In Fig. 12a, the antenna sections are each approximately a half wavelength long. The transmission-line elements are inserted between successive antenna sections to obtain proper phasing of the antenna currents. A common arrangement is to use transmission lines which are

(a)   (b)
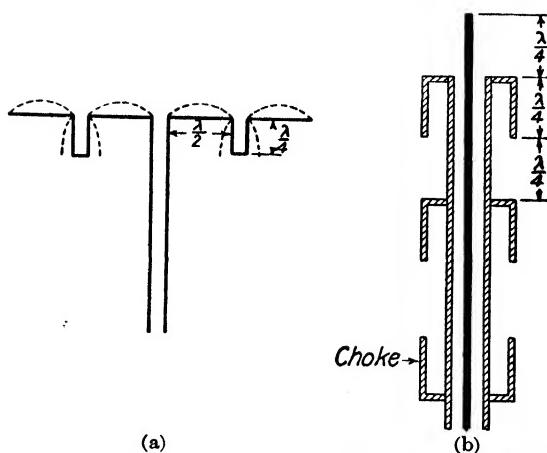
FIG. 12.—Colinear arrays.
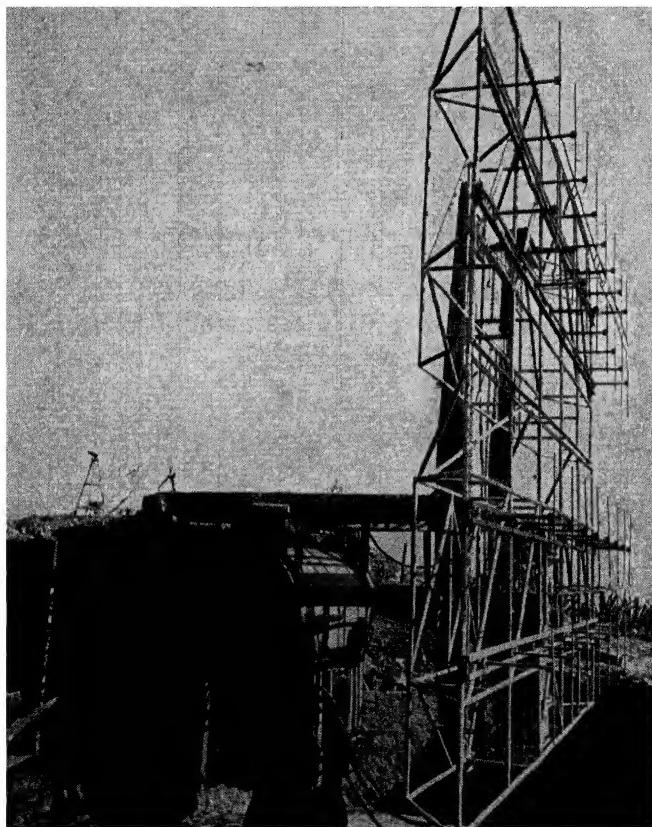
FIG. 13.—Rectangular array.

each a quarter wavelength long, in which case the antenna currents are in time phase. This results in a field distribution similar to that of the broadside array shown in Fig. 10a.

The array of Fig. 12b is fed by a coaxial transmission line. The first antenna of the array consists of the extended center conductor and the sleeve, forming a half-wavelength dipole antenna. Other antenna elements are formed by the sheath of the coaxial line and additional quarter-wavelength sleeves. The choke sleeve is a quarter wavelength long and serves to minimize induced currents in the coaxial-cable sheath below the array.

A rectangular array is sometimes used to obtain increased directivity. Figure 13 shows a rectangular array of half-wavelength dipole antennas. A metal screen behind the antennas serves as a reflector to give unidirectional radiation. The array shown is a broadside array, operating at a frequency of approximately 200 megacycles.

The turnstile array of Fig. 14 is an arrangement in which the two antennas on any one level cross at right

FIG. 14.—Turnstile antenna.

angles and have a 90-degree phase displacement of antenna currents. This results in a circularly symmetrical field pattern with horizontal polarization.
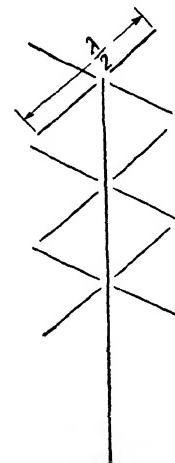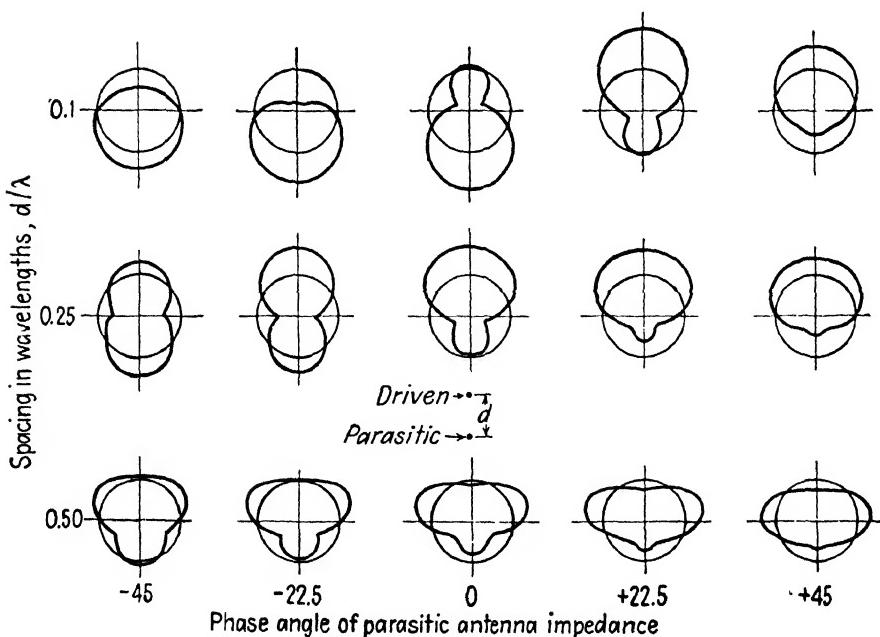
FIG. 15.—Field patterns in the horizontal plane for a driven antenna and a parasitic antenna.

By proper spacing of antenna elements and phasing of the antenna currents, the vertical radiation may be largely canceled.

19.07. Parasitic Antennas.—If a conducting wire is placed in the vicinity of a transmitting antenna and is oriented so as to be parallel to the electric-field lines, it becomes a *parasitic antenna*. Induced currents flow in the conductor and it absorbs and reradiates power according to its own directional radiation characteristics. The driven and parasitic antennas then comprise a two-element array with a field distribution which may be computed by the methods of the preceding section if the currents in the driven and parasitic antennas are known. The length of the parasitic antenna as well as the spacing between driven and parasitic antennas may be varied to obtain a variety of directivity patterns as shown in Fig. 15. The phase angles are those of the self-impedance of the parasitic antenna, which depends upon its length. The self-impedance of an isolated antenna which is less than a half wavelength long is capacitive, while that of an antenna greater than a half wavelength but less than a full wavelength long is inductive.

19.08. Loop Antennas.—Consider now the radiation from a circular loop antenna with dimensions which are small in comparison with the wavelength. Uniform current distribution in the loop is assumed and the radius of the loop is taken as a.

FIG. 16.—Loop antenna.

In the spherical coordinate system of Fig. 16, consider the vector potential at point P, with coordinates (r, θ, ϕ), resulting from a differential current element at Q, with coordinates (a, θ', ϕ'). The differential current element is $I a\, d\phi'$. The resultant vector potential at P is in the ϕ direction and, because of symmetry, the vector potential is uniform in the ϕ direction. In order to simplify the following discussion, we shall let $\phi = 0$. The differential current element at Q makes an angle of ϕ' with respect to the resultant vector potential at P, hence Eq. (15.02-14) may be written

$$A_\phi = \frac{\mu I}{4\pi} \int_0^{2\pi} \frac{e^{-j\beta r}}{r} a \cos \phi'\, d\phi' \tag{1}$$

Referring to Fig. 16, we may let $r = r_0 - a \cos \psi$ in the exponential term of Eq. (1) and $r = r_0$ in the denominator. It may be shown that $\cos \psi = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos (\phi - \phi')$. For our problem, we have
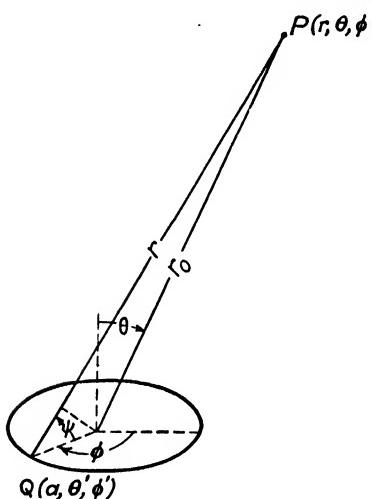
$\theta' = \pi/2$ and $\phi = 0$; therefore, $r = r_0 - a \sin \theta \cos \phi'$. Making this substitution in Eq. (1), we obtain

$$A_\phi = \frac{a\mu I e^{-j\beta r_0}}{4\pi r_0} \int_0^{2\pi} e^{j\beta a \sin \theta \cos \phi'} \cos \phi' \, d\phi' \qquad (2)$$

Taking the first two terms of the series expansion for the exponential, we have

$$A_\phi = \frac{a\mu I e^{-j\beta r_0}}{4\pi r_0} \int_0^{2\pi} (1 + j\beta a \sin \theta \cos \phi') \cos \phi' \, d\phi'$$

$$= \frac{j\pi I a^2 \mu}{2\lambda r_0} \sin \theta e^{-j\beta r_0} \qquad (3)$$

Inserting this value of A_ϕ into $\bar{H} = (1/\mu)\nabla \times \bar{A}$, we obtain the magnetic-intensity components,

$$H_r = \frac{j\pi I a^2}{\lambda r_0^2} \cos \theta e^{-j\beta r_0} \qquad (4)$$

$$H_\theta = \frac{\pi^2 I a^2}{\lambda^2 r_0} \sin \theta e^{-j\beta r_0} \qquad (5)$$

The electric intensity may be obtained by substituting $\bar{H}$ into $\nabla \times \bar{H} = j\omega\epsilon\bar{E}$. The scalar value of the radiation-field components are

$$H_\theta = \frac{\pi^2 I a^2}{\lambda^2 r_0} \sin \theta \qquad (6)$$

$$E_\phi = \eta H_\theta = \frac{\pi^2 I a^2 \eta}{\lambda^2 r_0} \sin \theta \qquad (7)$$

The radiation field of a loop antenna consists of a *TEM* spherical wave with intensities similar to those of the incremental antenna given in Eqs. (19.02-7 and 8), but with the electric and magnetic-intensity directions interchanged. The field pattern in the plane perpendicular to the loop is the same as that of the incremental antenna shown in Fig. 2.

The total radiated power and radiation resistance are:

$$P = \int_0^\pi \frac{\eta |H_\theta|^2}{2} 2\pi r_0^2 \sin \theta \, d\theta$$

$$= \frac{\pi \eta I^2}{12} \left(\frac{2\pi a}{\lambda}\right)^4 \qquad (8)$$

$$R_0 = \frac{2P}{I^2} = \frac{\pi \eta}{6} \left(\frac{2\pi a}{\lambda}\right)^4 = 31 \times 10^4 \left(\frac{a}{\lambda}\right)^4 \qquad (9)$$

As an example, a loop antenna with a value of $a/\lambda = 0.05$ has a radiation resistance of $R_0 = 1.94$ ohms, as compared with 73 ohms for the half-wavelength dipole antenna. The low radiation resistance of the loop antenna in comparison with its ohmic resistance results in relatively inefficient operation. Also, the low radiation resistance makes it difficult to obtain a proper impedance match to a transmission-line feeder.

Several modifications of the loop antenna have been developed in order to overcome the difficulty of low radiation resistance. The Alford loop,[1] shown in Fig. 17, has a current loop at the center of each side, with the currents in all four sides in phase. The nodal current points are brought
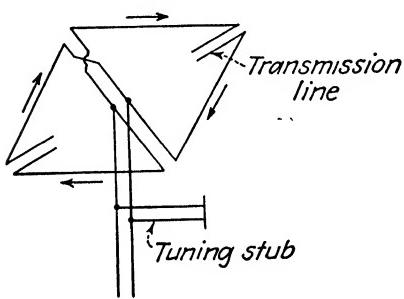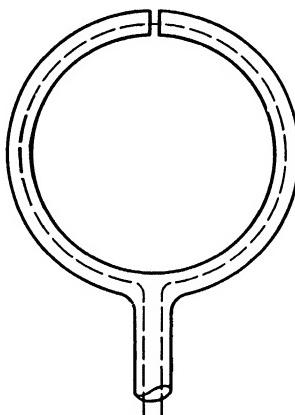
FIG. 17.—Alford loop.   FIG. 18.—Shielded loop antenna.

close together so that they do not radiate. This loop presents a relatively high reactive input impedance which is canceled by the reactance of the short-circuited stub. The radiation resistance of the Alford loop is $R_0 = 320 \sin^4 (\pi l/\lambda)$, where l is the length of one side. This may be made considerably greater than the radiation resistance of an ordinary loop antenna.

Figure 18 shows a loop antenna with an electrostatic shield. Antennas of this type are often used as transmitting or receiving antennas on airplanes. The electrostatic shield serves to minimize the static interference but does not appreciably alter the radiation characteristics as a loop antenna. Antenna currents flow on both the inside and outside surface of the outer conductor, the currents flowing on the outside being responsible for the radiation of power from the antenna.

19.09. Parabolic Reflectors.—The parabolic reflector is frequently used in microwave systems as a means of obtaining a high degree of directivity of transmitting and receiving antennas. Either a plane parabola or a

[1] ALFORD, A., and A. G. KANDOIAN, Ultra-high-frequency Loop Antennas, *Trans. A.I.E.E.*, vol. 59, p. 843; December, 1940.

paraboloid of revolution, such as shown in Fig. 19a or 19b, may be used. The antenna is placed at the focal point of the parabola. The following discussion applies principally to the paraboloid of revolution.
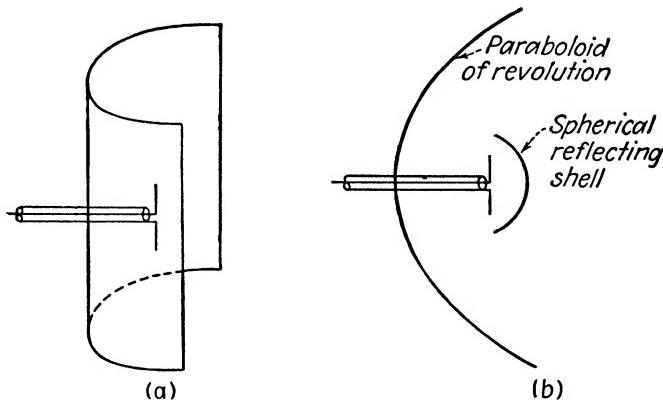
FIG. 19.—Parabolic antennas.

It is a well-known principle in optics that if a point light source is placed at the focal point of a parabolic mirror, the reflected rays emerge as parallel rays, concentrated in a narrow beam. Radio waves may be reflected by a parabolic reflector in much the same manner. The beam width is dependent
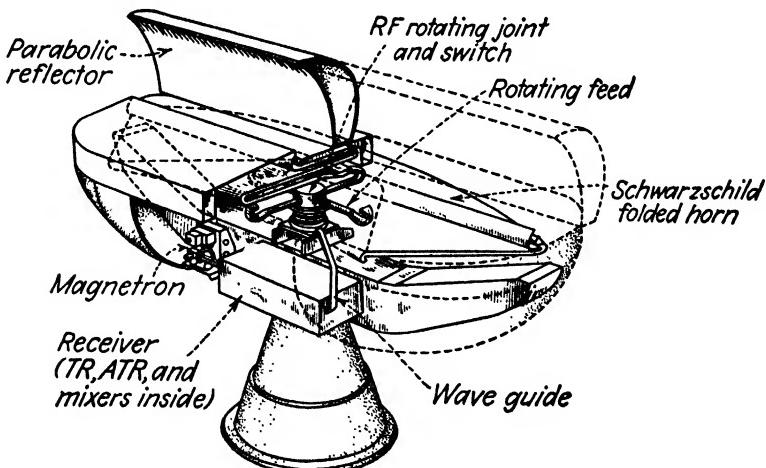
FIG. 20.—Schwarzschild antenna consisting of a plane parabola fed by a folded waveguide. Beam width is 0.6 degree in azimuth and 3 degrees in elevations. $\lambda = 3$ centimeters.

upon several factors. If the parabola is small, appreciable diffraction occurs at the edges of the parabola, resulting in increased beam width. In order to minimize diffraction the parabola should have a diameter exceeding 10 wavelengths.

Another factor tending to increase the beam width is the finite size of the antenna. Since the antenna is not a true point source, it produces focal defects, known as aberrations, which cause the rays to diverge from the ideal parallel beam. To minimize this difficulty, the parabola should have a focal distance of approximately one-quarter of its diameter.

A half-wavelength dipole antenna radiates uniformly in the horizontal direction but has a field pattern as shown in Fig. 5a in the vertical plane. Consequently, the parabola is uniformly illuminated in a horizontal plane but has most of the radiation concentrated at the center in the vertical plane. This results in a narrower beam in the horizontal direction than in the vertical direction. The direct radiation from the antenna which is not reflected by the parabola tends to spread out in all directions and hence partially destroys the directivity. The spherical reflecting shell, shown in Fig. 19b, or a parasitic antenna, may be used to direct all of the radiated energy toward the parabolic reflecting surface, thereby avoiding direct radiation.

PROBLEMS

1. An isolated linear antenna is 3λ long. The antenna is fed at a loop current point and the current is 10 amps.
   (a) Determine the angular positions of the maxima and minima of the radiation field in the vertical plane and sketch the field pattern.
   (b) Compute the electric intensity, magnetic intensity, and time-average power density at a point 10 km from the antenna in the direction of the maximum intensity.
   (c) Compute the radiation resistance.
2. Show that the time-average power radiated from an infinitesimal antenna, including the induction field terms, is the same as that of the radiation field alone. Obtain an expression for the peak value of reactive energy density stored in the field.
3. An array of m half-wavelength dipole antennas, such as that shown in Fig. 9, is mounted vertically over a perfectly conducting ground plane. Derive an expression for the electric intensity of the radiation field.
4. An end-fire array consists of six quarter-wavelength antennas mounted vertically over a perfectly conducting ground plane. The electric intensity in a horizontal plane at a distance of 25 miles from the antenna is to be 20 microvolts per m and the frequency is 50 megacycles per sec.
   (a) Specify the spacing between antennas for a maximum directivity of the radiated signal.
   (b) Sketch the field pattern in the horizontal and vertical planes.
   (c) What should be the antenna currents and phasing of the antenna currents?
   (d) Show how the antennas can be fed in order to obtain the desired phase relationships.
5. Derive an expression for the field pattern of the turnstile antenna of Fig. 14, assuming that all of the antennas in a vertical plane are fed in time phase and that the two antennas in any one horizontal plane have a phase displacement of 90°. Show that the resulting electric intensity is circularly polarized.

6. A loop antenna has a radius of 2 cm and a current of 100 ma. The frequency of the exciting source is 1,000 megacycles per sec.

(a) Compute the radiation resistance and power input, assuming that the antenna losses are negligible.

(b) Evaluate the electric intensity at a point in space 500 ft from the antenna and making an angle of 30° with respect to the plane of the antenna.

CHAPTER 20

IMPEDANCE OF ANTENNAS

The antenna must be considered not only from the viewpoint of a radiator or receiver of electromagnetic energy, but also as a circuit element in transmitting and receiving systems. For example, if maximum power transfer is to be realized, the networks feeding the antenna must be designed so as to present a conjugate impedance match at the antenna terminals. In certain applications, sharply tuned antennas are required, whereas in other applications, such as in television and pulse-modulated systems, the antennas must have a relatively constant input impedance over a wide range of frequencies. These few examples indicate the need for further knowledge of the subject of antenna impedances.

20.01. Input Impedance of Antennas.—The input impedance of an antenna is dependent upon its length, shape, the point where the antenna is fed, and the proximity of conductors or other objects which alter the field distribution in space.

A linear center-fed antenna, which is isolated in space, has an impedance variation with length (or impressed frequency) somewhat similar to that of the open-circuited transmission line shown in Fig. 10, Chap. 8, and has reactance characteristics similar to those shown in Fig. 6, Chap. 8. The length of the equivalent line is slightly greater than one-half of the antenna length. Resonant and antiresonant input impedances occur when the antenna length is approximately 5 per cent shorter than an integral number of half wavelengths, the deficiency being due to what is sometimes referred to as "end effects." The nature of these end effects will be discussed in connection with biconical antennas in the following chapter. The reactive component of input impedance alternates between capacitive and inductive reactance as the length of the antenna increases. The input impedance is a relatively low resistance at resonance and a high resistance at antiresonance.

In an array of antennas, the input impedance of any one antenna is dependent not only upon its self-impedance, but also upon the mutual impedance between the given antenna and all other antennas in the array. The use of a parabolic reflector or other reflecting device likewise alters the input impedance of an antenna.

A reciprocity theorem [1] of fundamental importance in the analysis of antennas states that if a voltage V, impressed at the terminals of one antenna, produces a current I in a second antenna, then the same voltage applied at the terminals of the second antenna will produce the same current
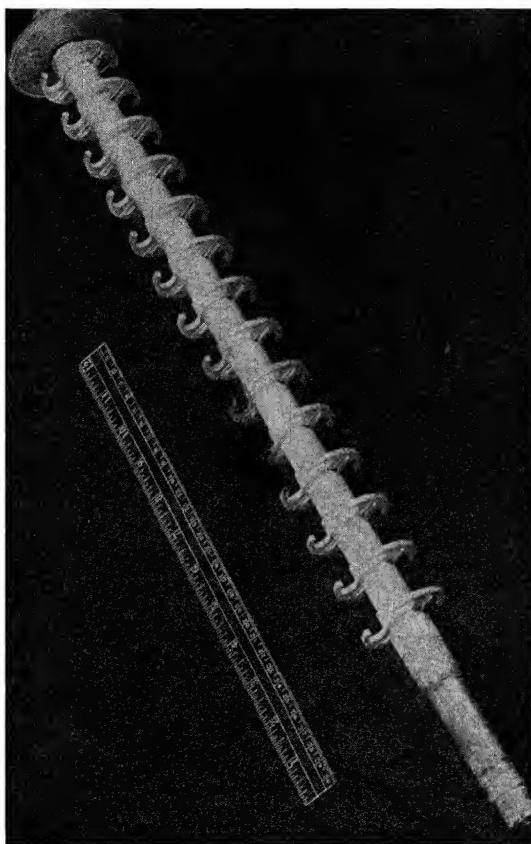
FIG. 1.—End fire array of curved dipole antennas. Beam width between half-power points is 28 degrees at $\lambda = 12$ centimeters. (*Courtesy of the M.I.T. Radiation Laboratory.*)

in the first antenna. Another way of stating this is that the transfer impedance of two antennas is the same with the generator connected to the terminals of either antenna, that is, $Z_t = V_1/I_2 = V_2/I_1$, where I_2 is the current in antenna 2 due to the voltage V_1 applied to antenna 1, and I_1 is the current in antenna 1 due to the voltage V_2 applied to antenna 2.

[1] CARSON, J. R., Reciprocal Theorems in Radio Communication, *Proc. I.R.E.*, vol. 17, p. 952; June, 1929.

As a result of the reciprocity theorem, the field pattern of an antenna or an array, as well as the self and mutual impedances of the various antenna elements, are the same when the antenna is used either as a transmitting or receiving antenna. The reciprocity theorem is restricted to isotropic mediums and does not apply if the field includes an ionized medium in the presence of a magnetic field, such as exists in the ionosphere region.

20.02. Methods of Evaluating Antenna Impedances.—There are four general methods of evaluating the input impedance of antennas. These are as follows:

1. If solutions of the wave equation are obtainable and the constants, representing the relative magnitudes of the various modes, can be evaluated, then the antenna current may be obtained for a given applied emf as a boundary-value problem. The input impedance is then computed as the ratio of voltage to current at the input terminals. Although this method represents a rigorous approach, the evaluation of the constants in the general solution of the wave equation involves considerable mathematical difficulty. For this reason, other simpler methods are used wherever possible.

2. The radiation resistance of an antenna may be determined by the Poynting-vector method described in the preceding chapter. In this method the field distribution is obtained in terms of either a known or an assumed current distribution in the antenna. The total radiated power is then computed by integrating the normal component of Poynting's vector over the surface of a large sphere having the antenna at its center. The radiation resistance is then obtained from $R_0 = 2P/I_0^2$, where P is the total radiated power and I_0 is the amplitude of the current at the position on the antenna where the current has its maximum value (the loop current point). For a perfectly conducting antenna, the radiation resistance would be the resistive component of input impedance, neglecting ohmic resistance, at the loop current point. This method does not give the reactive component of input impedance. When applied to arrays, this method may be used to obtain the total power radiated by the array, but it does not tell how much power is radiated by each individual antenna, nor does it enable us to evaluate the self and mutual impedances of the antenna elements in the array.

3. A modification of the Poynting-vector method is obtained if the surface over which Poynting's vector is integrated is chosen so as to coincide with the surface of the antenna conductor. This yields real and imaginary components of power, the real part representing power which is radiated from the antenna and the imaginary part representing energy which is stored in the induction field during one portion of the cycle and returned to the source during a later portion of the cycle. If ψ is the complex power, then the input impedance is $Z = 2\psi/I^2$, where I is the amplitude of the input current.

4. In the induced emf method, we again start with either a known or an assumed current distribution in the antenna and determine the resulting field distribution. The field induces a counter emf in the antenna which opposes the current in the antenna. Consequently the source must supply an equal and opposite emf to overcome the induced emf and thereby sustain the current flow in the antenna. The procedure, therefore, is to evaluate the induced emf by taking the line integral of electric intensity over the length of the antenna. The applied emf is equal and opposite to the induced emf and the ratio of applied emf to current, at the input terminals of the antenna, gives the input impedance.

We shall see presently that the third and fourth methods are essentially the same. Since these methods yield the most useful information, we shall consider them in greater detail in this chapter. In the foregoing discussion, no mention was made of the losses in the antenna itself. The power input to the antenna must equal the sum of the radiated power and the power loss in the antenna. The efficiency is the ratio of the radiated power to the power input. Since the losses in most antenna systems are small, they have very little effect upon the input impedance.

Fig. 2.—Center-fed antenna with a sinusoidal current distribution.

20.03. Field of a Linear Antenna—Exact Method.—In Sec. 19.03, the field distribution of a linear antenna was obtained by an approximate method in which only the radiation-field terms were included. Since the induced emf method requires a knowledge of the field distribution in the immediate vicinity of the antenna, we shall retrace our steps and derive more exact expressions for the field distribution.

Consider the center-fed linear antenna of Fig. 2, which is assumed to be isolated in space and to have a sinusoidal current distribution up to the feed point of the antenna. The current is given by

$$I = I_0 \sin \beta \left(\frac{l}{2} - Z \right) \qquad 0 < Z < \frac{l}{2}$$

$$I = I_0 \sin \beta \left(\frac{l}{2} + Z \right) \qquad -\frac{l}{2} < Z \leq 0$$

(1)

where the time function $e^{j\omega t}$ is assumed.

The charge distribution may be obtained by inserting the current from Eq. (1) into the equation of continuity, Eq. (3.02-7), which, for the one-dimensional case, may be written $\partial I/\partial Z = -\partial q_l/\partial t = -j\omega q_l$, where q_l is the charge per unit length of conductor. The charge distribution then becomes

$$q_l = -j\frac{I_0}{v_c}\cos\beta\left(\frac{l}{2} - Z\right) \qquad 0 < Z < \frac{l}{2}$$

$$q_l = j\frac{I_0}{v_c}\cos\beta\left(\frac{l}{2} + Z\right) \qquad -\frac{l}{2} < Z < 0 \tag{2}$$

The retarded scalar and vector potentials are obtained by inserting the current and charge from Eqs. (1) and (2) into (15.02-13 and 14), with the substitutions $q_r\,d\tau = q_l\,dZ$ and $J_c\,d\tau = I\,dZ$. The potentials at P in Fig. 2, resulting from the charge and current in the upper half of the antenna, then become

$$V = -j\frac{\eta I_0}{4\pi}\int_{Z=0}^{Z=l/2}\frac{e^{-j\beta r}}{r}\cos\beta\left(\frac{l}{2} - Z\right)dZ \tag{3}$$

$$A_z = \frac{\mu I_0}{4\pi}\int_{Z=0}^{Z=l/2}\frac{e^{-j\beta r}}{r}\sin\beta\left(\frac{l}{2} - Z\right)dZ \tag{4}$$

Similar expressions may be written for the potentials at P due to the current and charge in the lower half of the antenna.

The field intensities may be evaluated in terms of the potentials by using Eqs. (15.02-1 and 3),

$$\bar{H} = \frac{1}{\mu}\nabla\times\bar{A} \tag{15.02-1}$$

$$\bar{E} = -\nabla V - \frac{\partial\bar{A}}{\partial t} \tag{15.02-3}$$

When expressed in cylindrical coordinates for our particular problem, these become

$$E_z = -\frac{\partial V}{\partial z} - j\omega A_z \qquad E_\rho = -\frac{\partial V}{\partial \rho} \qquad E_\phi = 0 \tag{5}$$

$$H_\phi = -\frac{1}{\mu}\frac{\partial A_z}{\partial \rho} \tag{6}$$

The intensity component E_z at ρ resulting from the charge and current in the upper half of the antenna is obtained by inserting Eqs. (3) and (4) into the first of Eq. (5). Letting $f(r) = e^{-j\beta r}/r$, we have

$$E_z = j\frac{I_0}{4\pi}\left\{\eta\int_{Z=0}^{Z=l/2}\frac{\partial f(r)}{\partial z}\cos\beta\left(\frac{l}{2} - Z\right)dZ\right.$$

$$\left. - \omega\mu\int_{Z=0}^{Z=l/2}f(r)\sin\beta\left(\frac{l}{2} - Z\right)dZ\right\} \tag{7}$$

Referring to Fig. 1, we obtain

$$r^2 = \rho^2 + (z - Z)^2 \tag{8}$$

from which we obtain $\partial f(r)/\partial z = -\partial f(r)/\partial Z$. This substitution may be made in the first integral of Eq. (7) and the integration may then be completed by parts. We let $u = \cos \beta[(l/2) - Z]$ and $dv = [\partial f(r)/\partial Z] \, dZ$, and the first integral of Eq. (8) becomes

$$-\eta \int_{Z=0}^{Z=l/2} \frac{\partial f(r)}{\partial Z} \cos \beta \left(\frac{l}{2} - Z \right) dZ = -\eta f(r) \cos \beta \left(\frac{l}{2} - Z \right) \Big|_{Z=0}^{Z=l/2}$$

$$-\omega\mu \int_{Z=0}^{Z=l/2} f(r) \sin \beta \left(\frac{l}{2} - Z \right) dZ \tag{9}$$

When Eq. (9) is substituted into Eq. (8), the two integral terms cancel, leaving

$$E_z = -j \frac{\eta I_0}{4\pi} \frac{e^{-j\beta r}}{r} \cos \beta \left(\frac{l}{2} - Z \right) \Big|_{Z=0}^{Z=l/2}$$

Returning to Fig. 2, we find that when $Z = 0$ and $Z = l/2$ we have $r = r_0$ and $r = r_2$, respectively. The intensity E_z due to the current in the upper half of the antenna therefore becomes

$$E_z = -j \frac{\eta I_0}{4\pi} \left(\frac{e^{-j\beta r_2}}{r_2} - \frac{e^{-j\beta r_0}}{r_0} \cos \frac{\beta l}{2} \right) \tag{10}$$

A similar derivation, with the current and charge in the lower half of the antenna substituted into Eqs. (3) and (4), would yield the contribution to E_z resulting from the current and charge in the lower half of the antenna. Adding these two terms yields the total intensity,

$$E_z = j30I_0 \left(\frac{2e^{-j\beta r_0}}{r_0} \cos \frac{\beta l}{2} - \frac{e^{-j\beta r_1}}{r_1} - \frac{e^{-j\beta r_2}}{r_2} \right) \tag{11}$$

Expressions for E_ρ and H_ϕ may be obtained by inserting V and A_z into Eqs. (7) and (8), yielding, after considerable manipulation,

$$E_\rho = j \frac{30I_0}{\rho} \left\{ \frac{(z + l/2)}{r_1} e^{-j\beta r_1} + \frac{[z - (l/2)]}{r_2} e^{-j\beta r_2} - \frac{2z}{r_0} e^{-j\beta r_0} \cos \frac{\beta l}{2} \right\} \tag{12}$$

$$H_\phi = j \frac{I_0}{4\pi\rho} \left(e^{-j\beta r_1} + e^{-j\beta r_2} - 2e^{-j\beta r_0} \cos \frac{\beta l}{2} \right) \tag{13}$$

These equations may be used to evaluate the intensities for any value of ρ up to the surface of the antenna conductor.

20.04. Input Impedance of a Linear Antenna.—Having derived expressions for the intensities in the immediate vicinity of a linear antenna, we are now prepared to proceed with the impedance derivations. Assume that the linear antenna of Fig. 3 has a point source generator and a sinusoidal current distribution. The power density, flowing in a direction normal to the surface of the conductor, is given by Poynting's vector. The time-average power density may be expressed as $\psi = \frac{1}{2}E_z H_\phi^*$, where E_z and H_ϕ are the complex values of the intensities at the surface of the conductor and H_ϕ^* is the complex conjugate of H_ϕ. The power density represented by this expression is complex, the real part representing power lost by the system and the imaginary part representing reactive energy stored in the field. Integration of the power density over the surface of the conductor gives the total complex power

$$\psi = -\int_{-l/2}^{+l/2} (\tfrac{1}{2} E_z H_\phi^*) 2\pi a \, dz \qquad (1)$$

where a is the radius of the conductor and the negative sign denotes energy loss by the system.

The total power leaving the antenna may also be expressed in terms of the antenna current. Ampère's law $\oint \bar{H} \cdot d\bar{l} = I$, for our problem, becomes $2\pi a H_\phi = I$, or $H_\phi^* = I^*/2\pi a$. Making this substitution in Eq. (1), we obtain an alternate expression for the complex power

$$\psi = -\int_{-l/2}^{l/2} \tfrac{1}{2} E_z I_z^* \, dz \qquad (2)$$

Equations (1) and (2) offer two different interpretations for one and the same phenomenon. In Eq. (1), the complex power is expressed as the surface integral of Poynting's vector over a surface enclosing and coinciding with the antenna conductor surface. Our interpretation of Eq. (2) is that the intensity E_z opposes the current flow in the antenna, compelling the source to supply an equal and opposite electric intensity to sustain the current. The emf induced by the field is obtained by integrating E_z over the length of the antenna. This is equal and opposite to the applied emf.

The complex impedance at the input terminals is

$$Z = \frac{2\psi}{I_0^2} \tag{3}$$

where I_0 is the amplitude of the input current.

The intensities in the power expressions, Eqs. (1) and (2), are those at the surface of the antenna. These may be obtained from Eq. (20.03-11) by setting

$$r_0 = \sqrt{z^2 + a^2} \qquad r_1 = \sqrt{\left(\frac{l}{2} + z\right)^2 + a^2} \qquad r_2 = \sqrt{\left(\frac{l}{2} - z\right)^2 + a^2} \tag{4}$$

where a is the radius of the conductor. The value of E_z obtained in this manner is then inserted into Eq. (2) together with the current from the first of Eq. (20.03-1). The integration may be performed over the upper half of the antenna and the resulting power multiplied by two to obtain the total complex power. This gives

$$\psi = -j30I_0^2 \int_0^{l/2} \left[\frac{2e^{-j\beta\sqrt{z^2+a^2}}}{\sqrt{z^2+a^2}} \cos \frac{\beta l}{2} - \frac{e^{-j\beta\sqrt{[(l/2)+z]^2+a^2}}}{\sqrt{[(l/2)+z]^2+a^2}} \right.$$

$$\left. - \frac{e^{-j\beta\sqrt{[(l/2)-z]^2+a^2}}}{\sqrt{[(l/2)-z]^2+a^2}} \right] \sin \beta \left(\frac{l}{2} - z \right) dz \tag{5}$$

This integral may be evaluated by straightforward methods, although the actual manipulation is quite involved. The integration could be simplified by assuming that the conductor has zero radius, that is, $a = 0$. This would yield approximately the correct value of the real power and hence the resistive component of input impedance, as computed by inserting Eq. (5) into Eq. (3), would be the correct value. However, the imaginary component would be infinite, indicating an infinite reactance. In order to evaluate the reactive component of input impedance, it is therefore necessary to use the more exact expression. Schelkunoff [1] has obtained the following expressions for the resistance and reactance of the linear antenna by a method similar to that outlined above

$$R = 60(c + \ln \beta l - Ci\beta l) + 30(Si2\beta l - 2Si\beta l) \sin \beta l$$

$$+ 30 \left(c + \ln \frac{\beta l}{2} - 2Ci\beta l + Ci2\beta l \right) \cos \beta l \tag{6}$$

$$X = 60Si\beta l + 30(2Si\beta l - Si2\beta l) \cos \beta l$$

$$- 30 \left(\ln \frac{l\lambda}{2a^2} - 2.414 - Ci2\beta l + 2Ci\beta l \right) \sin \beta l \tag{7}$$

[1] Schelkunoff, S. A., "Electromagnetic Waves," pp. 369–374, D. Van Nostrand Company, Inc., New York, 1943.

where $c = 0.5572$ is Euler's constant and the cosine and sine integrals are as defined in Sec. 19.03. These terms may be readily evaluated for an antenna of any desired length if tables of the cosine and sine integrals are available. For example, if the antenna length is an odd integral number of half wavelengths long, we then have $\beta l = n\pi$ where n is an odd integer; then $\sin \beta l = 0$ and $\cos \beta l = -1$. Equations (6) and (7) then reduce to

$$R = 72.45 + 30 \ln n - 30Ci2\pi n \qquad (8)$$

$$X = 30Si2\pi n \qquad (9)$$

The resistive component of input impedance agrees with that obtained by the Poynting-vector method in Eq. (19.03-14). For a half-wavelength dipole antenna, we have $Ci2\pi n \approx 0$ and $Si2\pi n = 1.43$. The input impedance then becomes $Z = 72.45 + j43$ ohms. The reactive component of the input impedance may be made zero by using an antenna which is slightly less than a half wavelength long.

20.05. Validity of the Induced-emf Method.—Although the induced-emf method of evaluating antenna impedances yields results which agree favorably with measured values, this method embodies certain inconsistencies which lead us to question its validity. If we assume that the antenna is a perfect conductor, then the tangential intensity must be zero in order to satisfy the boundary conditions. This presents an embarrassing situation, since if E_z is zero at the conductor surface, then the normal component of Poynting's vector is likewise zero and there can be no power flow normal to the surface of the antenna. Does this mean that a perfectly conducting antenna could not radiate power? Experimental evidence indicates that the radiating properties of an antenna are actually improved as the conductivity of the antenna conductor increases. Intuitively we would expect that a perfectly conducting antenna would radiate just as effectively as an antenna with finite conductivity. If the radiated power does not leave the surface of the conductor, then where does it leave the antenna?

Before attempting an explanation of this anomaly, let us straighten out the matter of current distribution in the antenna. It is evident that, for a perfectly conducting antenna, the assumption of a sinusoidal current distribution is erroneous, since it yields a tangential intensity at the surface of the antenna which we know cannot exist. However, the value of E_z computed on the basis of an assumed sinusoidal current distribution is quite small and, for thin antennas, only a slight modification of the current distribution is necessary in order to cause E_z to vanish at the surface of the conductor, thereby satisfying the boundary conditions. Consequently, the current distribution along a thin, perfectly conducting antenna would differ slightly from a sinusoidal distribution, the discrepancy being most pronounced at the current nodes. As the conductor diameter approaches zero, the current distribution approaches a sinusoidal distribution.

In order to answer the question "where does the radiated power leave the antenna?" we must search for a surface over which the normal component of Poynting's vector is not zero. Could this be the extreme end surfaces of the antenna? In order to have a normal component of Poynting's vector, it would be necessary to have a tangential component of electric intensity at the end surfaces, but this is again ruled out by our assumption of a perfectly conducting antenna.

Let us now consider the antenna shown in Fig. 4, in which it is assumed that there is a finite separation distance between conductors at the feed point ab. An electric field exists in the region ab such that the line integral of electric intensity over this interval is equal to the impressed emf. A magnetic field also exists in this region owing to the current flow. Consequently we would expect to find a normal component of Poynting's vector over any surface enclosing the region ab. This leads us to suspect that the radiated power might depart from the antenna in the region between a and b. As the distance ab is decreased, the electric intensity increases in such a way that the line integral of electric intensity from a to b always equals the applied emf. In the foregoing derivations, a point generator was assumed, implying that the distance ab approaches zero. If this were true, however, the electric intensity would approach infinite value over the infinitesimal distance ab.

FIG. 4.—Field in the vicinity of the feed point of a dipole antenna.

Since the induced-emf method apparently yields reasonably correct results, we conclude that the complex power leaving a thin, perfectly conducting antenna may be determined by either of two methods.

1. Assume a sinusoidal current distribution and a point source feeding the antenna. Ignoring boundary conditions, evaluate E_z by the methods outlined above and insert this in either Eq. (20.04-1 or 2) to obtain the complex power flow.

2. With the true current distribution, E_z is zero at the surface of the antenna, but it is not zero in the region ab. We may therefore obtain the complex power flow by integrating Poynting's vector over a surface enclosing the region ab.

It is not clear why the two methods give essentially the same result. Apparently the slight modification in antenna current, which is necessary to satisfy the boundary conditions, causes a shift from a situation of distributed power flow from the conductor surface to one of concentrated power flow emanating from the region ab, without appreciably altering the value of the power or the complex impedance computed from it. Since all practical antennas have finite conductivity, the value of E_z is small but need not be zero.

20.06. Mutual Impedance of Linear Antennas.—When two or more antennas are located in a mutual field, a coupling exists between the antennas which influences the input impedance of each. Designating the self-impedances of a group of antennas at the input terminals as Z_{11}, Z_{22}, Z_{33}, etc., and the mutual impedances as Z_{12}, Z_{23}, etc., we may write the following familiar network equations

$$V_1 = I_1 Z_{11} + I_2 Z_{12} + \cdots I_n Z_{1n}$$

$$V_2 = I_1 Z_{21} + I_2 Z_{22} + \cdots I_n Z_{2n} \quad (1)$$

$$V_n = I_1 Z_{n1} + I_2 Z_{n2} + \cdots I_n Z_{nn}$$

where V_1, V_2, V_3, etc., are the applied emfs. From the reciprocity theorem, we find that the mutual impedances bear the relationship $Z_{12} = Z_{21}$, etc.

Consider the mutual impedance of two thin, parallel antennas of equal length as shown in Fig. 5. The current in antenna 1 produces a tangential component of electric intensity at the surface of antenna 2. This tangential electric intensity produces an emf in antenna 2 which, we will assume, opposes the current flow in antenna 2. It is then necessary for the source of antenna 2 to supply an equal and opposite emf to sustain the current flow in antenna 2. The complex power required of source 2 to overcome the induced emf is

$$\psi_2 = -\tfrac{1}{2} \int_{-l/2}^{l/2} E_{z12} I_2^* \, dz \qquad (2)$$

where E_{z12} is the intensity at the surface of antenna 2 due to the current in antenna 1, and I_2 is the current in antenna 2. The mutual impedance is then

$$Z_{12} = \frac{2\psi_2}{I_1 I_2} \qquad (3)$$

For linear antennas, Eq. (2) may be evaluated by assuming sinusoidally distributed currents in the two antennas. The electric intensity E_{z12} at the surface of antenna 2 resulting from the current in antenna 1 is then obtained from Eq. (20.03-11). The values of E_{z12} and I_2 are inserted into Eq. (2) and the integration is then performed to obtain ψ_2. Inserting this

Fig. 5.—Illustration for mutual impedance between antennas.
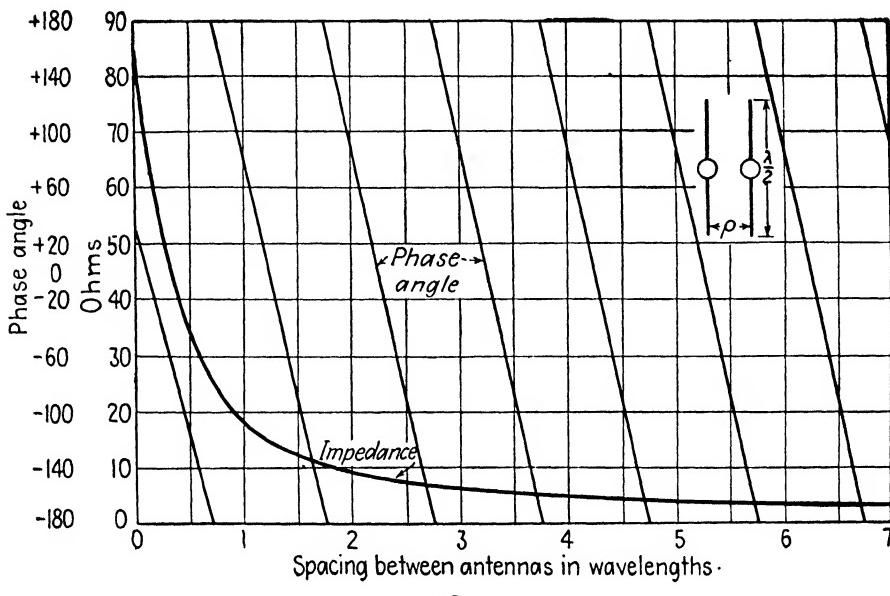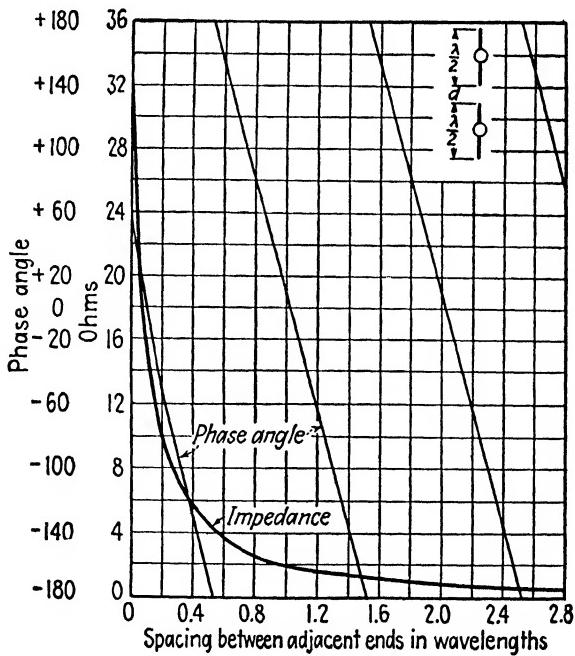
FIG. 6.—Magnitude and phase angle of the mutual impedance of two half-wavelength anten-nas; (a) parallel antennas, and (b) colinear antennas.

value of ψ_2 into Eq. (3) gives the mutual impedance. The integration is quite lengthy and will therefore be omitted. For two half-wavelength dipole antennas, the mutual impedance is

$$R_{12} = 30[2Ci\beta\rho - Ci\beta(r' + l) - Ci\beta(r' - l)]$$
$$X_{12} = 30[-2Si\beta\rho + Si\beta(r' + l) + Si\beta(r' - l)]$$

(4)

where ρ and r' are the distances shown in Fig. 5.

As an example, consider the mutual impedance resulting from two half-wavelength dipole antennas which are spaced a quarter wavelength apart. We then have $\beta\rho = \pi/2$. Referring to tables of Cix and Six in Jahnke and Emde, or Fig. 6, Chap. 19, we obtain

$$\beta\rho = \frac{\pi}{2} \qquad Ci\beta\rho = 0.47 \qquad Si\beta\rho = 1.37$$

$$\beta(r' + l) = 6.65 \qquad Ci\beta(r' + l) = 0 \qquad Si\beta(r' + l) = 1.43$$

$$\beta(r' - l) = 0.377 \qquad Ci\beta(r' - l) = -0.45 \qquad Si\beta(r') \qquad 0.37$$

We substitute these values into Eq. (4) and the mutual impedance is found to be $Z_{12} = 41.7 - j28.2$ ohms.

Graphs of the magnitude and phase angle of the mutual impedances of two half-wavelength antennas, arranged either as a parallel or colinear array, are shown in Figs. 6a and 6b.

PROBLEMS

1. Derive the expressions for E_z, E_ρ, and H_ϕ, given by Eqs. (20.03-11, 12, and 13), for the linear antenna.
2. Compute the input impedance of a center-fed, linear antenna which is 1 wavelength long. The frequency is 200 megacycles per sec.

CHAPTER 21

OTHER RADIATING SYSTEMS

A linear antenna, used either alone or in an array, constitutes a sharply tuned antenna, *i.e.*, relatively large impedance variations occur for small variations in frequency either side of the resonant or antiresonant frequency. Many applications require wide-band antennas which have a relatively small impedance variation over a wide range of frequencies. Certain types of antennas, such as the biconical antenna, the ellipsoidal antenna, the electromagnetic horn, and others, are inherently wide-band radiating and receiving systems. We shall consider several types of wide-band antennas in this chapter. The diffraction of electromagnetic waves passing through an aperture will also be treated.

21.01. Field of the Biconical Antenna. —The field in the vicinity of an antenna may be regarded as consisting of a superposition of a principal-mode field and higher spherical-mode fields such as those described in Sec. 15.11. Schelkunoff has shown that the biconical antenna has the unique feature that, insofar as the principal mode is concerned, it appears as an open-ended transmission line with uniformly distributed circuit parameters. Consequently this antenna offers an excellent opportunity to study radiation phenomenon from the viewpoint of transmission-line theory on the one hand, and the solutions of Maxwell's equations on the other.

Fig. 1.—The biconical antenna.

The biconical antenna of Fig. 1 is assumed to have two perfectly conducting cones, each of length l and cone angle $2\theta_1$. The electric field lines for the principal mode are arcs of circles terminating on charges on the surface of the cones as shown in Fig. 2a. The magnetic lines for the principal mode are circles concentric with the axis of the cone. If we assume a small cone angle, the principal-mode field may be assumed to be contained within the imaginary spherical surface s in Fig. 1. The electric field lines for two of the higher spherical modes are shown in Figs. 2b and 2c.

437
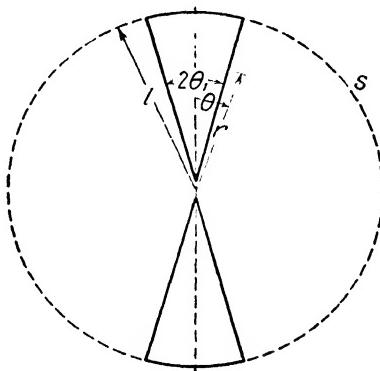
For the present, let us consider only the principal mode. The intensity components are E_θ and H_ϕ. The field is circularly symmetrical, hence $\partial/\partial\phi = 0$. The curl equations (13.06-1 and 2), expressed in spherical coordinates, then become

$$\frac{1}{r}\frac{\partial(rE_\theta)}{\partial r} = -j\omega\mu H_\phi \tag{1}$$

$$\frac{1}{r}\frac{\partial(rH_\phi)}{\partial r} = -j\omega\epsilon E_\theta \tag{2}$$

$$\frac{\partial}{\partial\theta}(\sin\theta H_\phi) = 0 \tag{3}$$

$$\frac{\partial}{\partial\phi}rE_\theta = 0 \tag{4}$$

where the time function $e^{j\omega t}$ is implied.
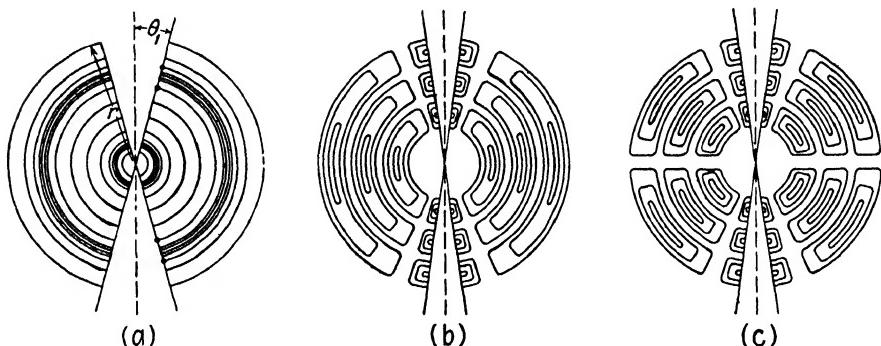
FIG. 2.—Field of the biconical antenna, (a) principal mode, (b) and (c) higher order modes.

An explicit expression for E_θ is obtained by multiplying Eq. (1) by r and then differentiating with respect to r. Substitution of $[\partial(rH_\phi)]/\partial r$ from Eq. (2) into this expression yields

$$\frac{\partial^2(rE_\theta)}{\partial r^2} = -\beta^2 rE_\theta \tag{5}$$

where $\beta = \omega\sqrt{\mu\epsilon}$.

Similarly, for the magnetic intensity we have

$$\frac{\partial^2(rH_\phi)}{\partial r^2} = -\beta^2 rH_\phi \tag{6}$$

Solutions of Eqs. (5) and (6) which also satisfy Eqs. (3) and (4) are

$$E_\theta = \frac{1}{r \sin \theta} (Ae^{-j\beta r} + Be^{j\beta r}) \tag{7}$$

$$H_\phi = \frac{1}{\eta r \sin \theta} (Ae^{-j\beta r} - Be^{j\beta r}) \tag{8}$$

where $\eta = \sqrt{\mu/\epsilon}$.

Equations (7) and (8) contain outgoing and reflected wave terms corresponding to the principal mode. The intensities decrease as $1/r$ due to the spherical nature of the waves. The ratio of electric to magnetic intensity for either the outgoing or reflected wave is equal to the intrinsic impedance of free space. Boundary conditions are satisfied since the electric intensity is normal to and the magnetic intensity tangential to the conducting surface.

The voltage between corresponding points on the two conductors is due to the field of the principal mode only.[1] This voltage may be computed as the line integral of electric intensity over the circular electric-flux path of the principal mode, or

$$V = -\int_{\theta_1}^{\pi-\theta_1} E_\theta r \, d\theta \tag{9}$$

Upon inserting E_θ from Eq. (7), we obtain

$$V = -(Ae^{-j\beta r} + Be^{j\beta r}) \int_{\theta_1}^{\pi-\theta_1} \frac{1}{\sin \theta} \, d\theta$$

$$= (Ae^{-j\beta r} + Be^{j\beta r})2 \ln \cot \frac{\theta_1}{2} \tag{10}$$

Ampère's law $\oint \vec{H} \cdot d\vec{l} = I$ may be used to evaluate the current. Integrating over a circular path concentric with the cone, we have $2\pi r \sin \theta H_\phi = I$, and inserting H_ϕ from Eq. (8) gives

$$I = \frac{2\pi}{\eta} (Ae^{-j\beta r} - Be^{j\beta r}) \tag{11}$$

From the viewpoint of a transmission line, the characteristic impedance Z_0 of the biconical antenna is the ratio of voltage to current for either the outgoing or reflected wave. Equations (10) and (11) give

$$Z_0 = \frac{\eta}{\pi} \ln \cot \frac{\theta_1}{2} \approx 120 \ln \frac{2}{\theta_1} \tag{12}$$

The second form of Eq. (12) is valid for small cone angles only.

[1] SCHELKUNOFF, S. A., "Electromagnetic Waves," p. 447, D. Van Nostrand Company, Inc., New York, 1943.

The characteristic impedance is independent of r; therefore, insofar as the principal mode is concerned, the biconical antenna has the properties of a uniform transmission line. The inductance and capacitance per unit length are also independent of r by virtue of the fact that the conductor diameter increases in direct proportion to r.

Although the biconical antenna appears as an open-ended transmission line, the energy of the outgoing wave is only partially reflected at the surface s, in Fig. 1, the remaining portion being transformed into higher-mode energy. Schelkunoff has shown that, from the viewpoint of the principal mode, this transfer of energy into the higher modes may be treated as a lumped impedance terminating a uniform transmission line.

21.02. Impedance of the Biconical Antenna.—The input impedance of the biconical antenna may be obtained from the transmission-line equation (8.06-4)

$$Z = Z_0 \left(\frac{Z_R + jZ_0 \tan \beta l}{Z_0 + jZ_R \tan \beta l} \right) \qquad (8.06\text{-}4)$$

where Z_R is the equivalent terminating impedance representing the effects of energy transfer from the principal mode to the higher modes.

Let us now consider, in a general way, the method used by Schelkunoff in determining the terminating impedance Z_R. The voltage difference between conductors (taken over a circular electric-flux path) is not affected by the presence of the higher-mode fields; hence we designate this voltage $V_0(r)$, where the zero subscript represents the principal mode. The current, however, contains components due to the principal mode as well as all higher-order modes. The current expressed in Eq. (21.01-11) represents the principal-mode current only. The higher-mode currents are represented by $I'(r) = I_1(r) + I_2(r) + \cdots I_n(r)$, where the subscripts denote the various modes. The total current at any point distant r from the apex is therefore

$$I(r) = I_0(r) + I'(r) \qquad (1)$$

where $I_0(r)$ is the principal mode current.

At the ends of the antenna $(r = l)$ the current must vanish; hence we have

$$I_0(l) = -I'(l) \qquad (2)$$

The equivalent terminating impedance is the ratio of voltage to current at the open end of the antenna, or

$$Z_R = \frac{V_0(l)}{I_0(l)} = -\frac{V_0(l)}{I'(l)} \qquad (3)$$

The problem of evaluating the terminating impedance, therefore, is to evaluate the higher-mode currents $I'(l)$ at the end of the antenna for a given impressed voltage. Schelkunoff accomplished this by considering the biconical antenna as consisting of two regions, (1) the region inside of the spherical surface s and (2) the region exterior to s. The distant field
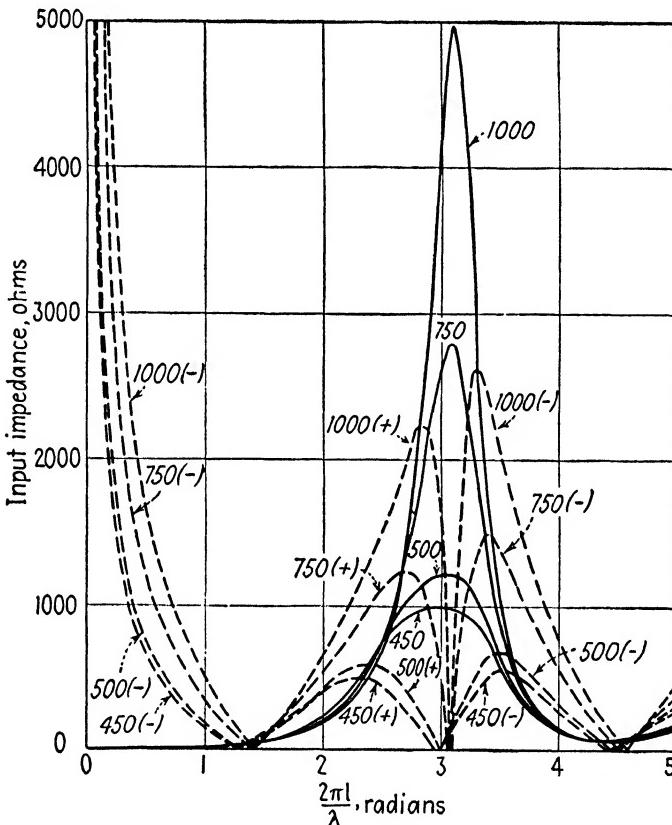
Fig. 3.—Input impedance of hollow biconical antennas for various values of characteristic impedance. Solid lines represent resistance, and dotted lines, reactance.

of the biconical antenna was then assumed to be the same as that of a linear antenna. This distant field was expressed in terms of spherical modes. The field inside of s was likewise expressed in terms of spherical modes, with the constants evaluated in such a manner as to satisfy the boundary conditions. The intensities corresponding to the field inside s and that outside s must have the same value at the boundary surface s. This requirement also makes it possible to evaluate constants in the spherical-mode solutions. Once the expressions for the field intensities have been obtained, the higher-mode currents can readily be evaluated as a

boundary-value problem; *i.e.*, the tangential magnetic intensities at the conductor surface must equal the surface-current densities for the various modes. The terminating impedance is then computed from Eq. (3) and this, in turn, is inserted into Eq. (8.06-4) to obtain the input impedance.

Figure 3 shows the variation of impedance with βl for biconical antennas having various values of characteristic impedance. The characteristic impedance for a given cone angle may be obtained from Eq. (21.01-12).
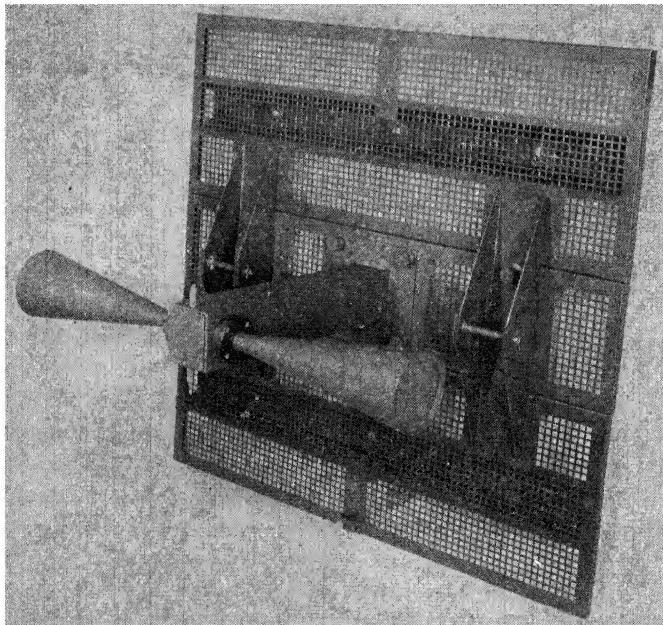
Fig. 4.—Biconical antenna for use in the frequency range from 132 to 156 megacycles. (*Courtesy of the Bell Telephone Laboratories.*)

The resonant length of the biconical antenna is somewhat shorter than a half wavelength. The resonant impedance is a pure resistance of approximately 72 ohms and is reasonably constant for a wide range of cone angles. As the cone angle increases, the antiresonant impedance decreases and the impedance variation with frequency is reduced. Consequently the biconical antenna with a large cone angle is essentially a broad-band antenna.

21.03. Higher-mode Fields of the Biconical Antenna.—The higher-mode fields of the biconical antenna are of the general form described in Sec. 17.14. The field is circularly symmetrical; hence we let $m = 0$ and use equations similar to Eqs. (17.14-14 or 15).

The radial component of electric (or magnetic) intensity may be represented by

$$E_r = \frac{A}{r} z_n(kr)[AP_n(\cos\theta) + BQ_n(\cos\theta)] \tag{1}$$

where $z_n(kr)$ is a general representation for any combination of spherical Bessel functions. In the region exterior to s, we use the spherical Hankel function $h_n^{(2)}(kr)$ defined by Eq. (15.08-1) and discard the second-kind Legendre function $Q_n(\cos\theta)$ because of its infinite value at $\theta = 0$ and $\theta = \pi$. The intensity in the region outside s then becomes

$$E_{r0} = \frac{A}{r} h_n^{(2)}(kr)P_n(\cos\theta) \tag{2}$$

Inside of the surface s the second-kind Bessel function is discarded because of its infinite value at $r = 0$, hence $z_n(kr) = j_n(kr)$. If we assume an infinitely thin antenna (zero cone angle), then $Q_n(\cos\theta)$ is discarded. The radial intensity inside s then becomes

$$E_{ri} = \frac{A}{r} j_n(kr)P_n(\cos\theta) \tag{3}$$

Expressions for $j_n(kr)$, $h_n^{(2)}(kr)$, and $P_n(\cos\theta)$ for the first few modes may be obtained from Eq. (15.08-1 and 2) and (15.11-15). The electric field lines for the first two higher-order modes are shown in Figs. 2b and 2c. The field in the region exterior to s is the same as that shown in Figs. 2b and 2c, except that the small loops terminating on the conductor in Fig. 2 are not present.

21.04. Other Wide-band Antennas.[1,2]—In general, the variation of impedance of an antenna with frequency decreases as the transverse dimensions of the antenna increase. Figure 4 shows a typical biconical antenna and Fig. 5 shows an antenna consisting of a cone and a disc. These are both wide-band antennas.

Schelkunoff has extended the method previously described for biconical antennas so as to apply to antennas of arbitrary size and shape. In this analysis, an average characteristic impedance is obtained for the particular antenna considered. The terminal impedance which is the equivalent of the higher-mode field effects is then obtained by the method described for the biconical antenna. The average characteristic impedance and equivalent terminal impedance are then substituted into an equation similar to

[1] SCHELKUNOFF, S. A., Theory of Antennas of Arbitrary Size and Shape, *Proc. I.R.E.*, vol. 29, pp. 493–521; September, 1941.

[2] MOULLIN, E. B., The Radiation Resistance of Surfaces of Revolution, Such as Cylinders, Spheres, and Cones, *J.I.E.E.*, vol. 88, p. 50; March, 1941; also discussions in *J.I.E.E.*, vol. 88, p. 171; June, 1941.
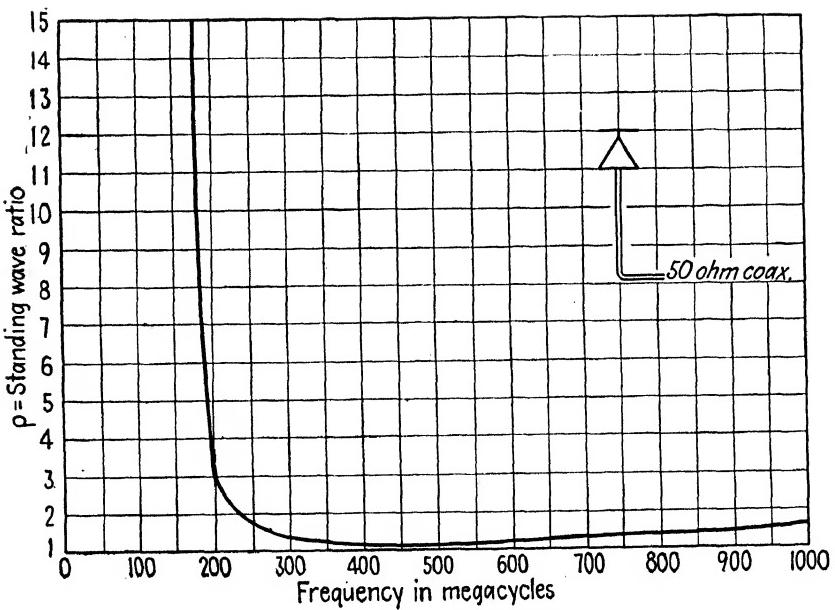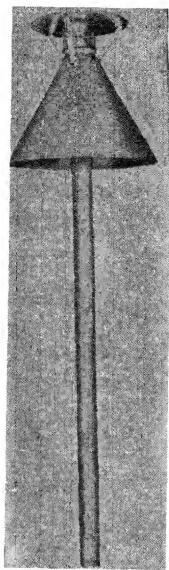
FIG. 5.—Discone antenna and standing-wave ratio. (*Courtesy of Federal Telecommunications Laboratories.*)
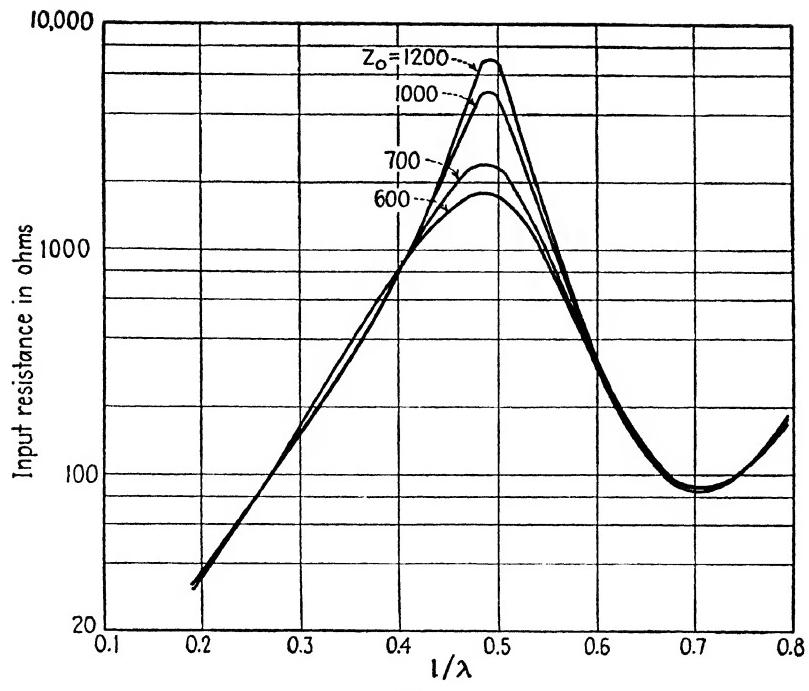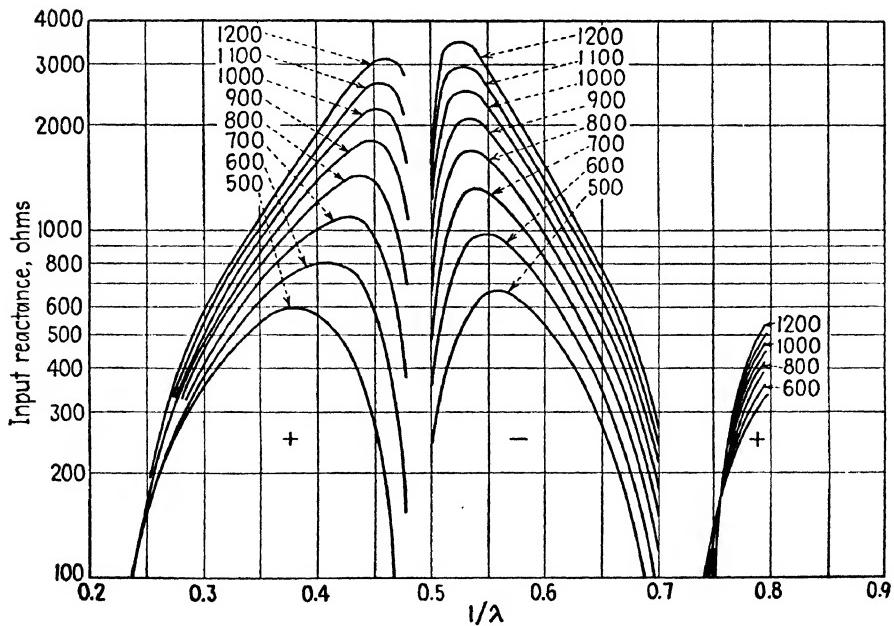
FIG. 6.—(a) Resistance and (b) reactance of hollow cylindrical antennas having various values of average characteristic impedance.

445

Eq. (8.06-4) (but derived for nonuniform lines) to obtain the input imped-
ance of the antenna.

The average characteristic impedance of a cylindrical antenna is

$$Z_0 = 120 \left(\ln \frac{l}{a} - 1 \right) \tag{1}$$

where l is the total length of the antenna and a is the radius. Figures 6a
and 6b show the resistive and reactive components of impedance as func-
tions of l/λ for cylindrical antennas
having various average character-
istic impedances.

21.05. The Sectoral Horn.—An
open-ended wave guide, which is
excited by a microwave source, will
radiate energy from the open end
of the guide into space. Since the
characteristic wave impedance in
the guide differs from the intrinsic
impedance of free space, the out-
going wave meets an abrupt dis-
continuity at the open end of the
guide. Consequently there is a re-
flected wave in the guide which, in
combination with the outgoing wave,
produces a standing wave. The
magnitude of the standing wave in
the guide is a measure of the degree
of impedance mismatch. At the
opening of the guide, diffraction of the radiated field causes the radiated
energy to scatter and results in poor directivity.

(a)

(b)

FIG. 7.—The sectoral horn.

However, if the wave guide is terminated by an electromagnetic horn,
such as shown in Fig. 7, there is a smooth impedance transformation from
the characteristic wave impedance of the guide to the intrinsic impedance
of free space. Consequently, the electromagnetic horn serves as an imped-
ance transfo mer, thereby increasing the radiated power and decreasing
the reflection in the guide. The horn also improves the directivity of the
field pattern. In many respects, the electromagnetic horn is similar to the
acoustical horn.

Consider the sectoral horn shown in Fig. 7, terminating a rectangular
guide.[1] The guide is assumed to be excited in the $TE_{0,n}$ mode. Cylindrical
coordinates are used, with the origin at the apex of the extended horn. For

[1] The analysis presented here follows that given by W. L. BARROW and L. J. CHU,
Theory of the Electromagnetic Horn, *Proc. I.R.E.*, vol. 27, pp. 51–64; January, 1939.

the $TE_{0,n}$ mode, we have $E_\rho = E_\phi = H_z = 0$ and the curl equations (13.06-1 and 2) become

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}(\rho H_\phi) - \frac{1}{\rho}\frac{\partial H_\rho}{\partial\phi} = j\omega\epsilon E_z \tag{1}$$

$$\frac{1}{\rho}\frac{\partial E_z}{\partial\phi} = -j\omega\mu H_\rho \tag{2}$$

$$\frac{\partial E_z}{\partial\rho} = j\omega\mu H_\phi \tag{3}$$

The solution of the wave equation in cylindrical coordinates, Eq. (15.05-14), for our particular problem, may be written

$$E_z = [C_1 J_\nu(k\rho) + C_2 N_\nu(k\rho)]\cos\nu\phi \tag{4}$$

The $TE_{0,n}$ modes have no intensity variation in the z direction. Only the $\cos\nu\phi$ terms are present if the guide is excited by a transverse antenna which is centered in the guide.

For convenience, the Bessel functions of the first and second kind may be replaced by Hankel functions of the second kind. Inserting E_z from Eq. (4) into (1) to (3), we obtain the magnetic intensities

$$E_z = A\cos\nu\phi\, H_\nu^{(2)}(k\rho) \tag{5}$$

$$H_\rho = \frac{A\nu}{j\omega\mu\rho}\cdot\sin\nu\phi H_\nu^{(2)}(k\rho) \tag{6}$$

$$H_\phi = -j\frac{A}{\eta}\cos\nu\phi\,\frac{\partial H_\nu^{(2)}(k\rho)}{\partial(k\rho)} \tag{7}$$

where $k = 2\pi/\lambda$. In general, the Bessel and Hankel functions in Eqs. (4) to (7) are of nonintegral order, hence ν is usually not an integer. In the horn the wave spreads as it travels outward and the propagation of the outgoing wave in the radial direction is governed by the second-kind Hankel function. The fact that nonintegral-order Hankel functions have infinite value for zero argument need not concern us here, since our analysis extends only to the junction of the guide and horn and therefore does not include the origin of the coordinate system.

To satisfy the boundary conditions, E_z must be zero at the side walls of the horn. This requires that $\cos\nu\phi_1 = 0$ where $2\phi_1$ is the flare angle of the horn. We therefore have

$$\nu\phi_1 = \frac{n\pi}{2} \qquad\text{or}\qquad \nu = \frac{n\pi}{2\phi_1} \tag{8}$$

where n is an odd integer denoting the half-wave periodicity in the ϕ direction.

In its impedance-transforming properties, the electromagnetic horn is similar to the tapered transmission line. In Sec. 10.12 the exponential line, terminated in its characteristic impedance, was shown to have an outgoing voltage wave given by $V = V_0 e^{-(\delta+j\beta)x}$, where $e^{-\delta x}$ represents the voltage transformation resulting from the taper. A wave function of this type is a solution of the differential equation $\partial V/\partial x = -(\delta + j\beta)V$.

We now assume that the horn has an exponential taper and that E_z satisfies the same differential equation as the voltage on the exponential line, or

$$\frac{\partial E_z}{\partial \rho} = -(\delta + j\beta)E_z \tag{9}$$

Inserting E_z from Eq. (5) into (9) gives

$$\frac{\delta + j\beta}{k} = -\frac{H_\nu^{(2)\prime}(k\rho)}{H_\nu^{(2)}(k\rho)} \tag{10}$$

where

$$H_\nu^{(2)\prime}(k\rho) = \frac{\partial H_\nu^{(2)}(k\rho)}{\partial(k\rho)}$$

Replacing the Hankel function by its value for small and large arguments, we obtain

$$H_\nu^{(2)}(k\rho) = \frac{1}{\nu!}\left(\frac{k\rho}{2}\right)^\nu + j\frac{(\nu-1)!}{\pi}\left(\frac{2}{k\rho}\right)^\nu \qquad k\rho \ll 1 \tag{11}$$

$$H_\nu^{(2)}(k\rho) = \sqrt{\frac{2}{\pi k\rho}}\, e^{j[-k\rho+(2\nu+1)\pi/4]} \qquad k\rho \gg 1 \tag{12}$$

Inserting these into Eq. (10), with the substitution $k = 2\pi/\lambda$, we obtain the transformation constant δ and phase constant β,

$$\left.\begin{aligned}
\delta &= \frac{n\pi}{2\rho\phi_1} \\[2ex]
\beta &= \frac{2\pi}{\lambda}\frac{\pi}{\{[(n\pi/2\phi_1)-1]!\}^2}\left(\frac{\pi\rho}{\lambda}\right)^{[(\pi n/\phi_1)-1]}
\end{aligned}\right\} \qquad k\rho \ll 1 \tag{13}$$

$$\left.\begin{aligned}
\delta &= \frac{1}{2\rho} \\[2ex]
\beta &= \frac{2\pi}{\lambda}
\end{aligned}\right\} \qquad k\rho \gg 1 \tag{14}$$

Small values of $(k\rho)$ correspond to the throat region of the horn, whereas large values correspond to the mouth region. Equation (14) assumes that

the horn is many wavelengths long. The phase and group velocities in the throat region of the horn are practically the same as those in the guide. At the mouth of the horn the transformation constant approaches zero, and the phase and group velocities approach the velocity of propagation in free space.

HORIZONTAL SECTION

TRANSVERSE SECTION

VERTICAL SECTION
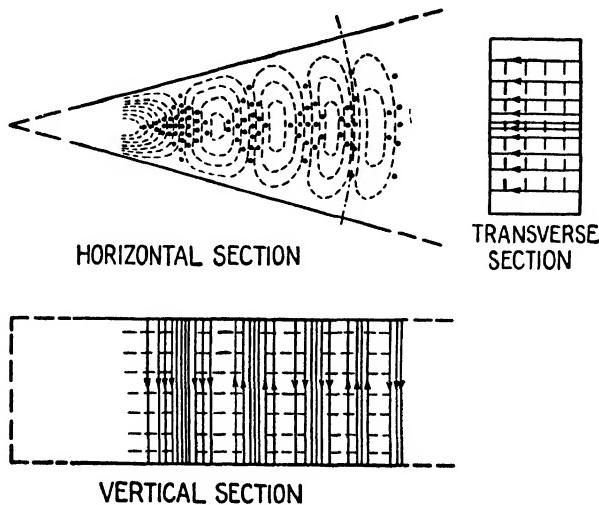
FIG. 8.—Field of the $TE_{0,1}$ mode in a sectoral horn.

The characteristic wave impedance is the ratio of the transverse components of electric and magnetic intensity, $Z_0 = E_z/H_\phi$. Inserting Eqs. (5) and (7) into this expression, we obtain

$$Z_0 = j\eta \frac{H_\nu^{(2)}(k\rho)}{H_\nu^{(2)\prime}(k\rho)} \tag{15}$$

and substituting Eq. (10) gives

$$Z_0 = -j\eta \left(\frac{k}{\delta + j\beta} \right)$$

$$= \frac{2\pi\eta}{\lambda} \left(\frac{\beta}{\delta^2 + \beta^2} - \frac{j\delta}{\delta^2 + \beta^2} \right) \tag{16}$$

At the throat, the characteristic wave impedance is approximately equal to that in the guide, whereas at the mouth, the characteristic wave impedance is practically equal to the intrinsic impedance of free space.

The electromagnetic field in the horn is gradually transformed from the *TE* wave of the wave guide to a *TEM* wave in free space, *i.e.*, the radial component of electric intensity E_ρ decreases as the wave progresses through the horn. Figure 8 shows the electric and magnetic field lines for a $TE_{0,n}$ wave in a sectoral horn.

21.06. Radiation Field of Electromagnetic Horns.—The radiation field of the electromagnetic horn may be obtained by applying either Huygens' principle or the equivalence principle. Later, we shall see how the equivalence principle may be applied to a somewhat simpler problem—the diffraction of a uniform plane wave in passing through a rectangular aperture. The analysis of the radiation field of the electromagnetic horn involves lengthy integrations and hence will not be considered here.[1]
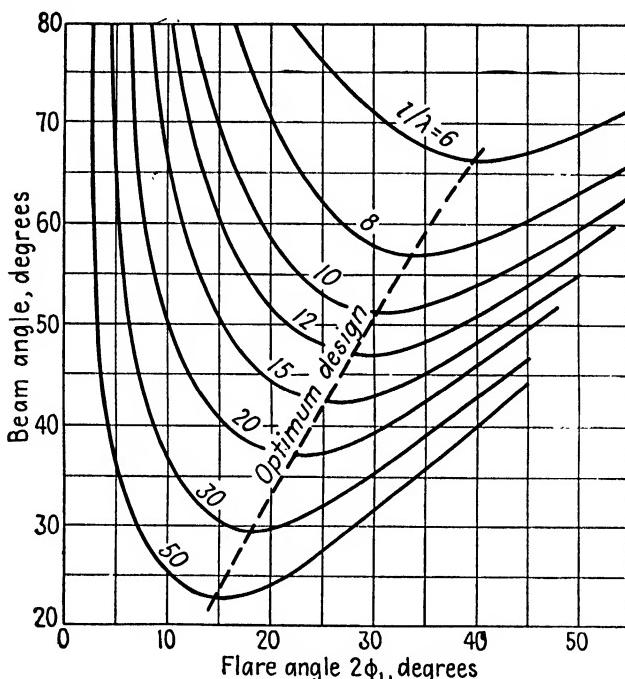
FIG. 9.—Beam angle between half-power points for sectoral horns having various lengths and flare angles.

Figure 9 shows the beam angle of sectoral horns as a function of flare angle and length. The optimum flare angle $2\phi_1$ is seen to vary between 40 degrees for a short horn to 15 degrees for a long horn.

The directional radiating properties of the electromagnetic horn result in a very appreciable power gain in the direction of maximum radiation. Figure 10 shows the power gain as compared with the radiation from a dipole antenna. The power is proportional to the square of the electric intensity.

The radiation field pattern of the sectoral horn in the plane perpendicular to the plane of the flare is less directional than in the plane of the flare.

[1] BARROW, W. L., and L. J. CHU, *loc. cit.*
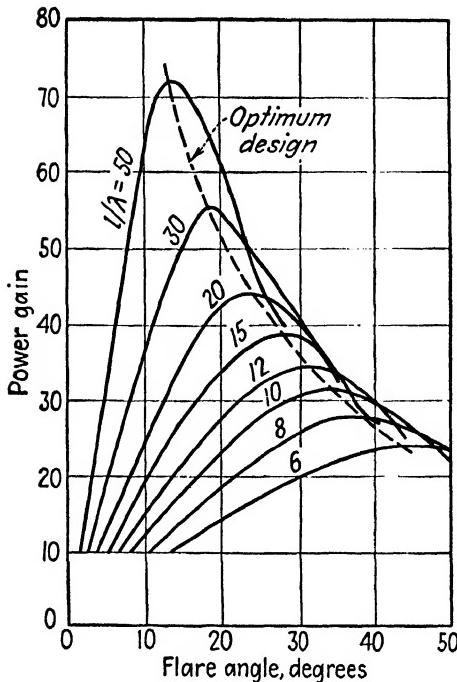
FIG. 10.—Power gain of a sectoral horn compared with a one-half wavelength antenna.

The over-all directivity may be improved by the use of the pyramid horn, shown in Fig. 11a, which is flared in both directions. The conical horn of Fig. 11b may be used with circular guides. Maximum directivity of long conical horns is obtained with a flare angle of approximately 40 degrees.

21.07. The Equivalence Principle.— Huygens showed that if the field distribution is known over any surface enclosing the source, such as surface *s* in Fig. 12, this may be used as the basis for determining the field distribution in space outside of *s*. According to this point of view, each point on the surface *s* is considered to be a secondary source of radiation, emitting spherical wavelets, the summation of which determines the field in space. Kirchhoff formulated this principle in terms of a scalar wave function, and presented a method of analysis which is commonly used
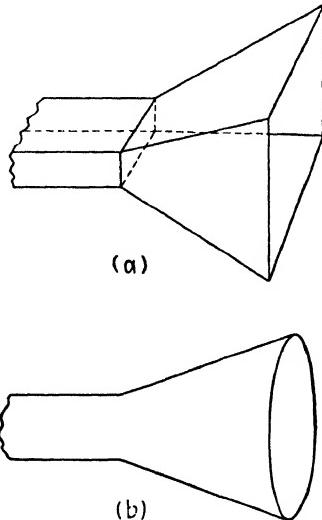
FIG. 11.—Pyramid and conical horns

in the solution of optical and acoustical diffraction problems. This method is also applicable to microwave diffraction problems, provided that the wavelength is small in comparison with the dimensions of the aperture.

The equivalence principle is a modification of the Huygens-Kirchhoff principle and is valid even though the aperture dimensions are comparable to the wavelength.[1] In the equivalence method, we again start with a known field distribution over the closed surface *s*. However, in this method we remove the source and replace it by an imaginary current sheet coinciding with surface *s*, the magnitude and phase of the currents being such as to produce the identical field distribution on *s* as existed with the original source. The field distribution in space exterior to *s* will then be the same either for the original source or for the imaginary current sheet and may be evaluated in terms of the currents in the equivalent current sheet.

Fig. 12.—Illustration of the equivalence principle.

In order to determine the value of the currents in the equivalent current sheet, we restore the original source and reverse the direction of the currents in the current sheet. With the proper value of currents (but reversed in direction) the fields exterior to *s* due to the original source and the current sheet are equal and opposite and hence cancel each other. Inside of the surface *s* the field is identical to that of the source alone, since the current sheet merely serves to terminate the field at *s*. The field intensities are then discontinuous across the boundary and the current density on *s* may be evaluated in terms of this discontinuity.

Our boundary conditions require that the discontinuity in the tangential component of magnetic intensity must equal the surface electric-current density on the equivalent current sheet. The tangential magnetic intensity H_t is perpendicular to the electric current density $\bar{J}$ and may be expressed by $\bar{J} = -\bar{n} \times \bar{H}_t$, where $\bar{n}$ is a unit vector in the direction of the outward drawn normal.

There is also a discontinuity in the tangential component of electric intensity at *s*. This is a violation of the boundary conditions of Sec. 13.08 and consequently it cannot exist physically. However, we may handle this case theoretically by inventing a fictitious magnetic current of density $\bar{M}$ on the surface *s*. The magnetic-current density serves to terminate the tangential electric intensity. Since the magnetic-current density is perpendicular to the tangential electric intensity $\bar{E}_t$, we write $\bar{M} = \bar{n} \times \bar{E}_t$.

The currents which we are seeking are those which produce the same field

[1] Schelkunoff, S. A., Some Equivalence Theorems and Their Application to Radiation Problems, *Bell System Tech. J.*, vol. 15, pp. 92–112; January, 1936.
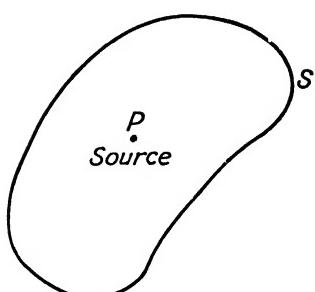
outside of s as the original source. These are equal and opposite to those given above, or

$$J = \bar{n} \times \bar{H}_t \tag{1}$$

$$\overline{M} = -\bar{n} \times \overline{E}_t \tag{2}$$

The current densities in the equivalent current sheet may therefore be evaluated in terms of the known values of electric and magnetic intensity on s and we may then proceed to evaluate the field outside of s in terms of these currents.

The curl equations may be written so as to include the hypothetical magnetic currents in a nondissipative medium as follows:

$$\nabla \times \bar{E} = -\overline{M} - \mu \frac{\partial \bar{H}}{\partial t} \tag{3}$$

$$\nabla \times \varPi = J + \epsilon \frac{\partial \bar{E}}{\partial t} \tag{4}$$

Two vector potentials are required, one in terms of the electric current and the other in terms of the magnetic current,

$$\bar{A} = \frac{\mu}{4\pi} \int_s \frac{J e^{-j\beta r}}{r} \, ds \tag{5}$$

$$\bar{F} = \frac{\epsilon}{4\pi} \int_s \frac{\overline{M} e^{-j\beta r}}{r} \, ds \tag{6}$$

In terms of the vector potentials, the intensities become

$$\bar{E} = -j\omega\bar{A} - \frac{1}{\epsilon} \nabla \times \bar{F} - j\frac{v_c^2}{\omega} \nabla(\nabla \cdot \bar{A}) \tag{7}$$

$$\varPi = -j\omega\bar{F} + \frac{1}{\mu} \nabla \times \bar{A} - j\frac{v_c^2}{\omega} \nabla(\nabla \cdot \bar{F}) \tag{8}$$

The procedure, therefore, is to evaluate the electric and magnetic currents in terms of the known intensities on s. The vector potentials are obtained by inserting the currents into Eqs. (5) and (6). Equations (7) and (8) are then used to evaluate the field intensities in space. The equivalence principle is particularly useful in finding the field distribution in various types of diffraction problems, or the radiation fields of open-ended coaxial lines and electromagnetic horns.

21.08. Diffraction of Uniform Plane Waves.—As an application of the equivalence principle, let us consider the diffraction of a uniform plane wave in passing through a rectangular aperture in a perfectly absorbing

screen as shown in Fig. 13. A perfectly absorbing screen would have an intrinsic impedance equal to that of free space and a very high attenuation constant, and would therefore absorb all of the incident energy except for that which passes through the aperture. The relationships derived here may be used as a first approximation for the diffraction of a uniform plane wave by an aperture in a conducting screen.

In optics, if the aperture plane *A* of Fig. 13 is uniformly illuminated and the viewing screen is placed close to the aperture, the image of the aperture
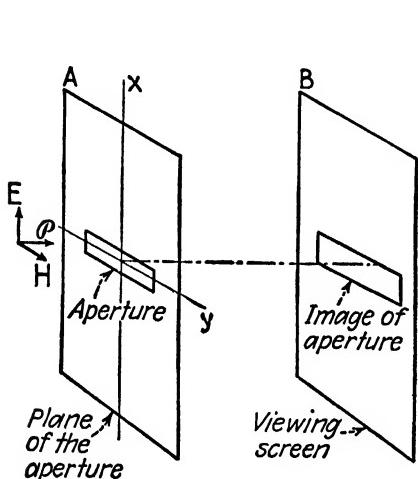
FIG. 13.—Example of plane-wave diffraction.
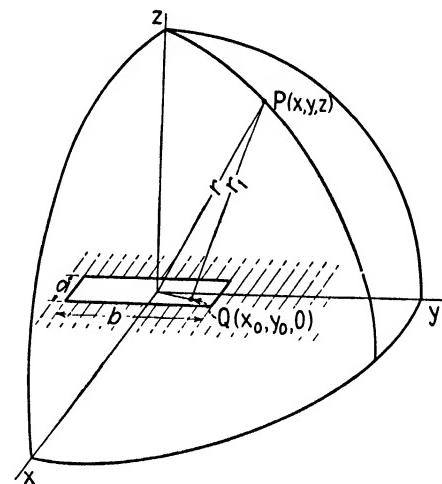
FIG. 14.—Coordinates for the diffraction of a uniform plane wave.

appears on the viewing screen. As the spacing between the two planes is increased, the image begins to show fuzzy variations in intensity at the edges. Upon further increasing the spacing, the image gradually loses its original shape and the intensity variations at the edges become more pronounced, with definitely distinguishable light and dark bands parallel to the edges of the image. The light bands result from reinforcement and the dark bands from cancellation of the waves. Fresnel's diffraction occurs for small spacings, while Fraunhofer diffraction corresponds to large spacings between the object and image planes, the mathematical solutions differing for the two cases.

Radio waves experience a similar diffraction phenomenon. Let us redraw Fig. 13 as shown in Fig. 14, with the plane of the aperture coinciding with the xy plane. Point $Q(x_0, y_0, 0)$ is in the plane of the aperture, whereas $P(x, y, z)$ is assumed to be in the viewing screen. The plane-wave source is assumed to be below the xy plane and the intensities E_{x0} and H_{y0} in the plane of the aperture are assumed to be uniform.

Since the intensities are uniform over the aperture surface, the currents in the equivalent current sheet are likewise uniform, and Eq. (21.07-1 and 2) give

$$J_{x0} = -H_{y0} \qquad M_{y0} = -E_{x0} \qquad (1)$$

Also, we have

$$\frac{E_{x0}}{H_{y0}} = \frac{M_{y0}}{J_{x0}} = \eta \qquad (2)$$

Inserting Eq. (1) into (21.07-5 and 6), we obtain the vector potentials

$$A_x = -\frac{\mu H_{y0}}{4\pi} \int \frac{e^{-j\beta r_1}}{r_1} \, ds \qquad (3)$$

$$F_y = -\frac{\epsilon E_{x0}}{4\pi} \int \frac{e^{-j\beta r_1}}{r_1} \, ds \qquad (4)$$

We also note that

$$\frac{A_x}{F_y} = \eta \qquad (5)$$

Assume that the distance r is large in comparison with the aperture dimensions and that it is also large in comparison with the wavelength. We also assume that the wave at point P is a transverse electromagnetic wave and hence has no z component of electric or magnetic intensity. With these assumptions, the last term in Eq. (21.07-7) varies as $1/r^2$ and hence is negligible for large distances r. Discarding this term, Eq. (21.07-7) becomes

$$\bar{E} = -j\omega\bar{A} - \frac{1}{\epsilon} \nabla \times \bar{F} \qquad (6)$$

The last term in Eq. (6) may be written

$$\frac{1}{\epsilon} \nabla \times \bar{F} = \frac{1}{\epsilon} \left[\frac{\partial F_y}{\partial x} \bar{k} - \frac{\partial F_y}{\partial z} \right] \bar{l}$$

Fig. 15.—Slot antenna for $\lambda = 3$ centimeters. Beam width to half-power points is 10 degrees.

The term $(\partial F_y/\partial x)\bar{k}$ represents a z component of electric intensity at P which we have assumed to be negligible; hence

$$\frac{1}{\epsilon} \nabla \times \bar{F} = -\frac{1}{\epsilon} \frac{\partial F_y}{\partial z} \bar{l}$$

Equation (6) may therefore be expressed as

$$E_x = -j\omega A_x + \frac{1}{\epsilon} \frac{\partial F_y}{\partial z} \qquad (7)$$

and with the substitution of Eq. (5)

$$E_x = \frac{1}{\epsilon} \frac{\partial F_y}{\partial z} - j\omega\eta F_y \tag{8}$$

The problem now is to evaluate F_y and $\partial F_y/\partial z$ in terms of the intensity E_{x0} at the aperture. To evaluate F_y, we return to Eq. (4) and make the substitution

$$r_1 = [(x - x_0)^2 + (y - y_0)^2 + z^2]^{\frac{1}{2}}$$

$$= [(x^2 + y^2 + z^2) - 2(xx_0 + yy_0) + x_0^2 + y_0^2]^{\frac{1}{2}} \tag{9}$$
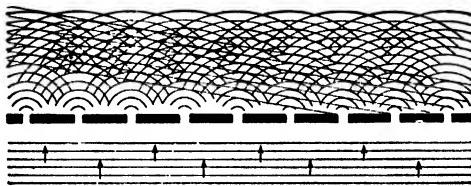
Fig. 16.—Diffraction of uniform plane waves by a grid.

Since x_0 and y_0 are small in comparison with x and y, the last two terms in Eq. (9) may be discarded. Expanding the remaining terms in a binomial series and retaining the first two terms of the series and letting

$$r^2 = x^2 + y^2 + z^2$$

we obtain

$$r_1 = r - \frac{xx_0 + yy_0}{r} \tag{10}$$

For conciseness, let $x/r = l$ and $y/r = m$, giving $r_1 = r - lx_0 - my_0$, which is substituted into Eq. (4), yielding

$$F_y = -\frac{\epsilon E_{x0} e^{-j\beta r}}{4\pi r} \int_{-a/2}^{a/2} e^{j\beta lx_0}\, dx_0 \int_{-b/2}^{b/2} e^{j\beta my_0}\, dy_0 \tag{11}$$

where we have treated $r_1 \approx r$ as a constant in the denominator of Eq. (4).

Equation (11) integrates to

$$F_y = \frac{\epsilon E_{x0} abe^{-j\beta r}}{4\pi r} \left[\frac{\sin (\pi la/\lambda)}{(\pi la/\lambda)} \frac{\sin (\pi mb/\lambda)}{(\pi mb/\lambda)} \right] \tag{12}$$

which we write

$$F_y = \frac{ke^{-j\beta r}}{r} \tag{13}$$

where

$$k = \frac{\epsilon E_{x0} ab}{4\pi} \left[\frac{\sin (\pi la/\lambda)}{(\pi la/\lambda)} \frac{\sin (\pi mb/\lambda)}{(\pi mb/\lambda)} \right]$$

We then have

$$\frac{\partial F_y}{\partial z} = -\frac{z}{r}\left(j\beta + \frac{1}{r}\right)\frac{k}{r}e^{-j\beta r}$$

$$= -\frac{z}{r}\left(j\beta + \frac{1}{r}\right)F_y \qquad (14)$$
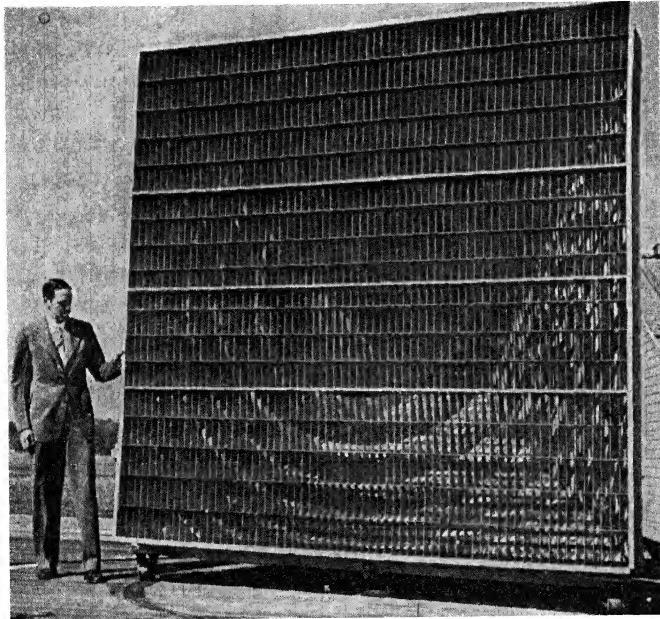
Fig. 17.—Metal lens antenna fed by a horn. Power gain over isotropic radiator is 12,000. Power gain is 1 per cent of maximum value at 1.8 degree off the axis of the beam. Wavelength, 7.5 centimeters.

Inserting Eqs. (13) and (14) into (8), together with the value of k, gives the desired expression for the electric intensity as

$$E_x = -\left\{\left[\frac{z}{r^2}\left(\frac{1}{r} + j\beta\right) + \frac{j\beta}{r}\right]\frac{E_{x0}ab}{4\pi}\right\}\frac{\sin(\pi la/\lambda)}{(\pi la/\lambda)}\frac{\sin(\pi mb/\lambda)}{(\pi mb/\lambda)}$$

$$= A\frac{\sin u}{u}\frac{\sin v}{v} \qquad (15)$$

where A is the bracketed term in Eq. (15), $u = \pi la/\lambda$ and $v = \pi mb/\lambda$.

The values of $\sin u/u$ and $\sin v/v$ determine, to a large extent, the variation of intensity in the viewing plane, since u and v vary with x and y, respectively, where x and y are the coordinates of the point P. This is apparent if we write $u = \pi ax/\lambda r$ and $v = \pi by/\lambda r$. The function $\sin u/u$ is

plotted in Fig. 2, Chap. 12.   From this graph, it is clear that nodal values of electric and magnetic intensity occur for values of u (or v) equal to π, 2π, etc.   Maximum intensities occur approximately when $u = 3\pi/2$, $5\pi/2$, etc.   This accounts for the light and dark bands found in the Fraunhofer diffraction pattern.

21.09. Optics and Microwaves.—We have found that microwaves behave, in many respects, like light waves, obeying the same laws of wave propagation, reflection, refraction, and diffraction.   Intuitively, we would expect this, since fundamentally they are one and the same physical phenomenon.   However, the electrical properties of materials, *i.e.*, the ϵ, μ, and σ parameters, may differ greatly in passing from microwaves to light waves.   These parameters are inextricably dependent upon the behavior of atoms and electrons in high-frequency electromagnetic fields.   Such phenomena as electron resonance within the atom or energy-level transitions seriously affect the values of these parameters.   If we were to pursue the subject further it would be necessary to make a fundamental analysis of the electrical properties of matter in high-frequency fields.

It is interesting to speculate on the possibilities of using various types of optical systems for microwaves.   Thus, if low-loss refractive materials are available, it may be possible to construct lenses, prisms, diffraction gratings, and other optical systems for operation at microwave frequencies.   In theory, at least, all of these are possible, although the physical size would probably be larger than the corresponding optical counterpart.

APPENDIX I

SYSTEMS OF UNITS

In the formulation of an absolute system of units, each quantity must be expressed in terms of a minimum number of basic, independent units. The choice of the basic units is somewhat arbitrary although they are usually chosen in such a manner that permanent, reproducible standards can be constructed. The units of mass, length, and time have been generally accepted as three of the basic units. In order to define electromagnetic quantities, one additional basic unit is required.

The *cgs electrostatic* and *electromagnetic systems of units* use the centimeter, gram, and second as the units of mass, length, and time, respectively. In the electrostatic system of units, the permittivity of free space ϵ_0 is taken as unity and the units and dimensions of all other quantities are dimensioned in terms of mass, length, time, and ϵ_0. In the electromagnetic system, the permeability of free space μ_0 is taken as unity, and mass, length, time, and μ_0 become the basic units. The quantities ϵ_0 and μ_0, however, are not independent, since Maxwell's equations show that the quantity

$$v_c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$$

is equal to the velocity of light, which is a measurable physical quantity ($v_c = 2.99796 \times 10^{10}$ centimeters per second). This establishes a relationship between the electrostatic and electromagnetic systems of units and accounts for the fact that the velocity of light appears in the conversion of quantities from either system of units to the other system.

Many of the electromagnetic equations contain both electric and magnetic quantities in a single mathematical expression. The question then arises as to whether the quantities in these equations should be expressed in (1) the electrostatic system of units, (2) the electromagnetic system of units, or (3) a mixed system of units in which the electrical quantities are expressed in electrostatic units while the magnetic quantities are in electromagnetic units. The latter system is commonly used and is known as the *Gaussian system of units*. This is a hybrid system of units, since both electrostatic and electromagnetic units may appear in the same equation; hence the velocity of light often appears in these equations as a proportionality constant.

From a practical viewpoint, the units of the electrostatic and electromagnetic systems proved to be of inconvenient size. Consequently, a practical system evolved in which the units are related to those in the electromagnetic system by powers of ten. The electrical quantities in the practical system of units are the familiar volt, ampere, coulomb, ohm, watt, joule, henry, farad, etc.

Giorgi showed that this practical system of units could be made into a self-consistent, absolute system of units by a suitable choice of the basic units. In this system of units, the meter, kilogram, and second are chosen as the units of mass, length, and time, respectively. This is known as the *absolute mks system of units*. The choice of the fourth basic unit is arbitrary. In recent years the unit of charge, the coulomb, has become generally accepted as the fourth basic unit. This choice is due largely to the fact that dimensional formulas of electric and magnetic quantities do not contain fractional powers when expressed in terms of mass, length, time, and charge. For example, if resistance were taken as the fourth basic unit, the dimensional formula for electric potential would be $[V] = [M^{1/2}LT^{-3/2}R^{1/2}]$, whereas, in terms of charge, the dimensions of potential become

$$[V] = [ML^2T^{-2}Q^{-1}]$$

Any system of units may be either rationalized or unrationalized. In unrationalized units, the unit of electric flux is chosen in such a manner that 4π lines of electric flux emanate from each unit of charge. This choice results in a factor of 4π appearing in the principal electromagnetic field equations; thus, in unrationalized units, Eqs. (13.01-1 to 4) become

$$\oint \bar{E} \cdot d\bar{l} = -\frac{\partial \phi}{\partial t}$$

$$\oint \bar{H} \cdot d\bar{l} = 4\pi i$$

$$\oint_s \bar{D} \cdot d\bar{s} = 4\pi q$$

$$\oint_s \bar{B} \cdot d\bar{s} = 0$$

In the rationalized system of units, the units are defined in such a manner that each unit of charge has one unit of electric flux emanating from it. This choice of units eliminates the 4π factor in the above equations, as well as in many other important electromagnetic field equations, thereby simplifying the writing of these equations. In the rationalized system of units, the 4π factor is incorporated into the permittivity and permeability, giving

$\epsilon_0 = 1/(36\pi \times 10^9)$ and $\mu_0 = 4\pi \times 10^{-7}$. The more common quantities, such as potential, charge, current, resistance, power, energy, capacitance, inductance, electric intensity, magnetic flux density, etc., are the same in rationalized and unrationalized units. However, the units of electric flux and electric flux density, as well as those of magnetic intensity and magnetomotive force, differ by a factor of 4π in rationalized and unrationalized units. It should be noted that, although rationalization removes the 4π factor from many equations, this factor reappears in other equations, such as Coulomb's law, Eq. (2.02-2), and the potential equations expressed in terms of charge and current (15.02-11 and 12). The rationalized mks system of units is used throughout this text.

CONVERSION TABLE

CONVERSION TABLE FOR RATIONALIZED MKS, ESU, AND EMU SYSTEMS OF UNITS. EQUALITY SIGNS ARE IMPLIED BETWEEN THE ELEMENTS OF A ROW, *i.e.*, 1 KILOGRAM = 1,000 GRAMS

Quantity	Symbol	Rationalized mks	Electrostatic	Electromagnetic
Length...................	l	1 meter	100 centimeters	100 centimeters
Mass........	m	1 kilogram	1,000 grams	1,000 grams
Time...................	t	1 second	1 second	1 second
Force.............. ..	f	1 newton	10^5 dynes	10^5 dynes
Work or energy..........	w, W	1 joule	10^7 ergs	10^7 ergs
Power.................	p, P	1 watt	10^7 ergs per second	10^7 ergs per second
Charge............... .	q	1 coulomb	3×10^9	10^{-1}
Charge density (surface)...	q_s	1 coulomb per meter^2	3×10^5	10^{-5}
Charge density (volume)...	q_r	1 coulomb per meter^3	3×10^3	10^{-7}
Current.................	i, I	1 ampere	3×10^9	10^{-1}
Current density....... ...	J	1 ampere per meter^2	3×10^5	10^{-5}
Electric potential.........	$V, \mathbf{E}$	1 volt	$\frac{1}{3} \times 10^{-2}$	10^8
Electric intensity.........	E	1 volt per meter	$\frac{1}{3} \times 10^{-4}$	10^6
Electric flux..............	ψ	1 coulomb	$12\pi \times 10^9$	$4\pi \times 10^{-1}$
Electric flux density......	D	1 coulomb per meter^2	$12\pi \times 10^5$	$4\pi \times 10^{-5}$
Conductivity............	σ	1 mho per meter	9×10^9	10^{-11}
Resistance..............	R	1 ohm	$\frac{1}{9} \times 10^{-11}$	10^9
Capacitance..............	C	1 farad	9×10^{11}	10^{-9}
Magnetomotive force......	$\mathcal{F}$	1 ampere turn	$12\pi \times 10^9$	$4\pi \times 10^{-1}$ gilbert
Magnetic intensity......	H	1 ampere turn per meter	$12\pi \times 10^7$	$4\pi \times 10^{-3}$ oersted
Magnetic flux...........	ϕ	1 weber	$\frac{1}{3} \times 10^{-2}$	10^8 maxwells
Magnetic flux density......	B	1 weber per meter^2	$\frac{1}{3} \times 10^{-6}$	10^4 gausses
Magnetic potential (vector)	$\overline{A}$	1 weber per meter	$\frac{1}{3} \times 10^{-4}$	10^6
Inductance..............	L	1 henry	$\frac{1}{9} \times 10^{-11}$	10^9
Permittivity of free space..	ϵ_0	$\frac{1}{36\pi} \times 10^{-9} = 8.85 \times 10^{-12}$ farad per meter	1	$\frac{1}{9} \times 10^{-20}$
Permeability of free space..	μ_0	$4\pi \times 10^{-7}$ henry per meter	$\frac{1}{9} \times 10^{-20}$	1

APPENDIX II

ELECTRICAL PROPERTIES OF MATERIALS

CONDUCTIVITY OF METALS

Material	Conductivity σ, mhos per meter at 20°C
Aluminum, commercial hard drawn.........	3.54×10^7
Brass (30% zinc).......................	$1.2\text{--}1.5 \times 10^7$
Constantin (60% cu, 40% ni).............	0.2×10^7
Copper, annealed.......................	5.80×10^7
Copper, hard drawn.....................	5.65×10^7
Iron (pure)............................	1.00×10^7
Manganin (84% cu, 12% mn, 4% ni)......	0.23×10^7
Nichrome..............................	0.10×10^7
Nickel................................	1.28×10^7
Silver (pure)....	6.14×10^7
Tin...............	0.860×10^7
Tungsten........	1.81×10^7

PROPERTIES OF SEMICONDUCTORS

Material	Conductivity σ, mhos per meter	Relative permittivity ϵ_r (dielectric constant)
Sea water....	3–5	80
Lake water......... ..	$10^{-2}\text{--}10^{-3}$	80
Distilled water........	2×10^{-4}	81
Wet earth....... . . .	$10^{-2}\text{--}10^{-3}$	5–15
Dry earth..............	$10^{-4}\text{--}10^{-5}$	2–5

Properties of Dielectrics at Low Frequencies

Material	Conductivity σ, mhos per meter	Relative permittivity (dielectric constant)
Bakelite	10^{-8}–10^{-10}	5.5
Glass	10^{-12}	5.4–9.9
Mica	10^{-11}–10^{-15}	5.6–6.0
Petroleum	10^{-14}	2.13
Porcelain	3×10^{-13}	5.7
Quartz (fused)	$<2 \times 10^{-17}$	4.5–5

Properties of Dielectrics at Microwave Frequencies [1-4]

Material	Loss tangent ($\approx$ power factor)	Relative permittivity (dielectric constant)	Wavelength, centimeters
Bakelite	0.012–0.093	3.98–4.58	10
Glass	0.0016–0.0089	3.84–6.70	10
Lucite	0.0068–0.0084	2.53–2.60	10
Mycalex	0.0030–0.0033	5.74–5.91	22.5
Nujol	0.0008	2.14	10
Polyethylene	0.0002–0.0012	2.25	6
Polystyrene	0.0005–0.002	2.53	6
Paraffin	0.0002	2.25	6
Quartz (fused)	0.0001	3.80	6
Rubber	0.0041–0.0059	2.69–2.77	10
Styralloy	0.00105	2.50	10
Water (distilled)	0.15	4	10
Wood	0.060–0.075	1.75–2.05	6

[1] The electrical properties of most materials at microwave frequencies vary appreciably, depending upon the composition of the material, humidity, temperature, etc.

[2] ENGLUND, C. R., Dielectric Constants and Power Factors at Centimeter Wavelengths, *Bell System Tech. J.*, vol. 23, pp. 114–129; January, 1944.

[3] ROBERTS, S., and A. VON HIPPEL, A New Method for Measuring Dielectric Constant and Loss in the Range of Centimeter Waves, *J. Applied Phys.*, vol. 17, pp. 610–616; July, 1946.

[4] "Microwave Transmission Data," Sperry Gyroscope Co., New York, May, 1944.

APPENDIX III

Formulas of Vector Analysis

	Rectangular coordinates	Cylindrical coordinates	Spherical coordinates
Conversion from rectangular coordinates		$x = \rho\cos\phi \quad y = \rho\sin\phi \quad z = z$	$x = r\cos\phi\sin\theta \quad y = r\sin\phi\sin\theta \quad z = r\cos\theta$
Gradient	$\nabla V = \dfrac{\partial V}{\partial x}\bar{i} + \dfrac{\partial V}{\partial y}\bar{j} + \dfrac{\partial V}{\partial z}\bar{k}$	$\nabla V = \dfrac{\partial V}{\partial \rho}\bar{\rho} + \dfrac{1}{\rho}\dfrac{\partial V}{\partial \phi}\bar{\phi} + \dfrac{\partial V}{\partial z}\bar{k}$	$\nabla V = \dfrac{\partial V}{\partial r}\bar{r} + \dfrac{1}{r}\dfrac{\partial V}{\partial \theta}\bar{\theta} + \dfrac{1}{r\sin\theta}\dfrac{\partial V}{\partial \phi}\bar{\phi}$
Divergence	$\nabla\cdot\bar{A} = \dfrac{\partial A_x}{\partial x} + \dfrac{\partial A_y}{\partial y} + \dfrac{\partial A_z}{\partial z}$	$\nabla\cdot\bar{A} = \dfrac{1}{\rho}\dfrac{\partial(\rho A_\rho)}{\partial \rho} + \dfrac{1}{\rho}\dfrac{\partial A_\phi}{\partial \phi} + \dfrac{\partial A_z}{\partial z}$	$\nabla\cdot\bar{A} = \dfrac{1}{r^2}\dfrac{\partial(r^2 A_r)}{\partial r} + \dfrac{1}{r\sin\theta}\dfrac{\partial A_\phi}{\partial \phi} + \dfrac{1}{r\sin\theta}\dfrac{\partial(\sin\theta\,A_\theta)}{\partial \theta}$
Curl	$\nabla\times\bar{A} = \begin{vmatrix} \bar{i} & \bar{j} & \bar{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix}$	$\nabla\times\bar{A} = \begin{vmatrix} \dfrac{1}{\rho}\bar{\rho} & \bar{\phi} & \dfrac{1}{\rho}\bar{k} \\ \dfrac{\partial}{\partial \rho} & \dfrac{\partial}{\partial \phi} & \dfrac{\partial}{\partial z} \\ A_\rho & \rho A_\phi & A_z \end{vmatrix}$	$\nabla\times\bar{A} = \begin{vmatrix} \dfrac{\bar{r}}{r^2\sin\theta} & \dfrac{\bar{\theta}}{r\sin\theta} & \dfrac{\bar{\phi}}{r} \\ \dfrac{\partial}{\partial r} & \dfrac{\partial}{\partial \theta} & \dfrac{\partial}{\partial \phi} \\ A_r & rA_\theta & r\sin\theta A_\phi \end{vmatrix}$
Laplacian	$\nabla^2 V = \dfrac{\partial^2 V}{\partial x^2} + \dfrac{\partial^2 V}{\partial y^2} + \dfrac{\partial^2 V}{\partial z^2}$	$\nabla^2 V = \dfrac{1}{\rho}\dfrac{\partial}{\partial \rho}\left(\rho\dfrac{\partial V}{\partial \rho}\right) + \dfrac{1}{\rho^2}\dfrac{\partial^2 V}{\partial \phi^2} + \dfrac{\partial^2 V}{\partial z^2}$	$\nabla^2 V = \dfrac{1}{r^2}\dfrac{\partial}{\partial r}\left(r^2\dfrac{\partial V}{\partial r}\right) + \dfrac{1}{r^2\sin\theta}\dfrac{\partial}{\partial \theta}\left(\sin\theta\dfrac{\partial V}{\partial \theta}\right) + \dfrac{1}{r^2\sin^2\theta}\dfrac{\partial^2 V}{\partial \phi^2}$

Index